Learning the Macro-Dynamics of U.S. Treasury Yields With Aribitrage-free Term Structure Models*

Marco Giacoletti Kristoffer T. Laursen Kenneth J. Singleton †

March 23, 2014

Abstract

*We are grateful for feedback from

†Marco Giacoletti is at the Graduate School of Business, Stanford University, Giacoletti_Marco@gsb.stanford.edu; Kristoffer T. Laursen is at the Graduate School of Business, Stanford University, klaursen@stanford.edu; and Kenneth J. Singleton is at the Stanford University, Graduate School of Business, and NBER, kenneths@stanford.edu

1 Introduction

Looking back over the past several decades, investors in government bond markets have faced several major financial crises,[1] unforeseen major changes in the operating procedures and transparency policies of the Federal Reserve,[2] and considerable lack of clarity about the future path and impact of fiscal policies. Within these economic and informational environments, it seems non-controversial to characterize bond-market participants as facing nontrivial learning problems as they price bonds, assess required compensations for bearing relevant factor risks, and forecast the future course of the term structure of interest rates.

This paper estimates an arbitrage-free dynamic term structure model (*DTSM*) in which the marginal investor in U.S. Treasury bonds exhibits *ex ante* learning about the joint distribution of the Treasury yield curve and the macroeconomy. The estimated learning rule has formal roots in the literature on Bayesian learning, and it reveals which aspects of risk and the state of the economy agents learn about and which aspects they (evidently) know from information embedded in the current yield curve. Versions of our learning-based pricing model replicate quite accurately the entire term structure of median survey forecasts of bond yields over the past twenty-five years. Moreover, our learning model reveals interesting patterns in how the "consensus investor" adjusted her forecasts of bond yields when (as we now know with hindsight) the Federal Reserve changed its operating procedures and disclosure policies.

Learning is introduced through the updating of the parameters governing a bond-market-specific stochastic discount factor $\mathcal{M}$ and the joint dynamics of the yield curve and other conditioning variables as new information arrives. $\mathcal{M}$ is parameterized with low-order principal components of yields (PCs) as the risk factors, thereby capturing the impact of the diverse bond-market relevant shocks in an empirically reasonable and parsimonious way.[3] In our most flexible *DTSMs* the conditioning information Z is higher dimensional, encompassing aspects of the macroeconomy and heterogeneity of investors' views. Consistent with recent history, the learning rule we posit accommodates unforeseen bond-market-relevant structural changes that are not immediately fully understood by market participants.

We embark on our exploration of the nature of learning in bond markets in Section 4 by endowing a consensus investor with the following learning rule: each month she computes maximum likelihood estimates of a three-factor Gaussian affine term structure model, normalized as in Joslin, Singleton, and Zhu (2011) (JSZ) with the first three PCs of yields ($\mathcal{P}$) as risk factors, and allowing for the possibility of down weighting of past data. Then forecasts of future bond yields are computed presuming that the estimated parameters of this *DTSM* are those of the true data-generating process and that these parameters will remain fixed over the forecast horizon. This *DTSM*-based learning rule produces forecast errors for yields

[1]These include the Asian currency crisis, Russian default and nearly concurrent failure of Long-Term Capital Management, the housing section collapse in the US, and the recent sovereign debt crisis in Europe.

[2]Beyond the policy experiment during the early 1980's, the Federal Reserve has instituted several unconventional policies and these decisions have affected how market professionals forecast interest rates (see, e.g., Swanson (2006), Wright (2011), and Krishnamurthy and Vissing-Jorgensen (2011)).

[3]Our characterization of the priced risks in bond markets in terms of PCs goes back at least to Litterman and Scheinkman (1991), and many subsequent studies have noted that their model-implied risk factors resemble (linear combinations of) PCs. Recently Joslin, Singleton, and Zhu (2011) have formalized the idea that *DTSMs* can be rotated so that the risk factors are PCs of yields.

that are strikingly similar to the errors of the median forecaster for the Blue Chip Financial Forecasters (BCFF) survey, across the entire yield curve and over the past twenty five years. This is true even though our basic model with learning conditions only on the information in the yield curve. Moreover, we show that forecasts that outperform the median BCFF forecaster are obtained by a *DTSM*-based learner who down weights historical data with a half-life of about five years. This down weighting, as we show formally, arises naturally in a setting where the consensus agent is concerned about structural change.

If we expand the conditioning to allow risk premiums to also depend on unspanned inflation and output risks as in Joslin, Priebsch, and Singleton (2013) (JPS), then *DTSM*-based rules systematically give smaller out-of-sample, root-mean-square forecast errors than median survey forecasts. The gain in forecasting performance of the JPS-based learning rule is especially large during the early 2000's in the run-up to the recent financial crisis. This learning rule also out performs the median professional during the crisis, when professionals were systematically, and incorrectly, expecting a rise in rates that was not realized. The discipline of the *DTSM*-based learning rules kept forecasts closer in line with future realized yields.

Though these *DTSM*-based learning rules are sophisticated in their dependence on arbitrage-free pricing models, they are also quite naive in their treatment of parameter drift.[4] Specifically, to determine the prices of bonds today using the fitted *DTSM*, the agent computes forecasts of future values of $\mathcal{M}$ holding the parameters fixed (equivalently, holding the parameters governing the risk-neutral distribution of $\mathcal{P}$ fixed). Additionally, with prices in hand, forecasts under the historical distribution are also computed assuming fixed parameters. Both behaviors are dynamically inconsistent in the face of parameter drift associated with learning.

Why then are the naive rules so accurate in matching the forecasts of the median professional? The answer is that, given the historical behavior of U.S. Treasury bond yields, *DTSM*-based learning is (along one key dimension) dynamically consistent after all. We show in Section 4 that , even though the consensus investor is allowed to update all of the parameters of her *DTSM* over time, there is very limited drift in the parameters ($\Theta^{\mathbb{Q}}$) governing the risk-neutral conditional mean of $\mathcal{P}$. This implies that the loadings across maturities that describe how bond yields are related to $\mathcal{P}$ are essentially time-invariant and, as a consequence, the consensus investor is in fact computing the optimal risk-neutral forecasts for pricing throughout our sample period. Only the forecasts under the historical distribution ($\mathbb{P}$) are inefficient. Moreover we show that, once $\Theta^{\mathbb{Q}}$ is fixed, our investor is effectively following an adaptive least-squares learning algorithm.[5]

With these results as background, in Section 6 we explore the properties of risk premiums and the nature of the parameter drifts implied by learning. Though the forecast errors from the *DTSM*- and survey-based learning rules are quite similar, we find that the implications of these rules for risk premiums–expected excess returns– in Treasury markets are very different. The implied premia are most different shortly after NBER recessions when the slope of the Treasury curve is near a peak steepness. We show that, during these periods, the discipline

[4]Nevertheless, because of the computational tractability of rules similar to this one, they are often adopted in studies of learning (see Cogley and Sargent (2008)).

[5]As is shown subsequently, this result is distinct from the observation in JSZ that, in unconstrained Gaussian *DTSMs*, maximum likelihood estimates of the historical $\mathbb{P}$-mean of $\mathcal{P}$ are identical to the least-squares estimates. In our setting there are constraints on the market prices of $\mathcal{P}$ risks.

of our *DTSM*-based learning rules leads to substantially more accurate forecasts of excess returns than those embedded in the professional surveys, particularly for long-term bonds.

The paper builds upon a growing literature that explores learning or survey expectations within *DTSMs*. One of the first papers to introduce survey expectations into the estimation of *DTSMs* was Kim and Orphanides (2012). Piazzesi, Salomao, and Schneider (2013) extend their framework by having survey forecasts represent subjective views that are distinct from those of the econometrician. Both of these studies presume that agents know the parameters underlying their structure of beliefs– there is no learning. Furthermore, up to measurement errors, their models presume that survey expectations are fully spanned by the first three *PC*s of bond yields. In contrast, we focus on learning through the lens of *DTSMs* and use these rules to characterize the learning implicit in the survey forecasts. For our data, large portions of the variation in survey forecasts are unspanned by the yield *PC*s, and this has material implications for interpreting risk premiums in the presence of learning.

Laubach, Tetlow, and Williams (2007) and Dewachter and Lyrio (2008) study learning within three-factor Gaussian *DTSMs* in which the factors are inflation, a measure of real activity, and the one-period bond yield. As documented in Joslin, Le, and Singleton (2013), when the risk factors are measures of output growth, inflation, and a bond yield, *DTSMs* tend to substantially misprice bonds, in their case mispricing of the ten-year bond exceeded 100 basis points for extended periods of time.[6] For the same reasons, we suspect that similar issues would arise with the vector-autoregression-based learning rules examined by Cieslak and Povala (2014), where they to imbed their rule in a *DTSM*. Large pricing errors are avoided in our learning environment by accommodating unspanned macroeconomic information that has predictive power for bond risk premiums.

Section 2 introduces our formal learning environments and connects *DTSM*-based learning to Bayesian learning. The data and *DTSMs* used in our empirical work are described in Section 3. Full Bayesian learning is compared to the relatively naive *DTSM*-based rules in Section 5. Section 7 takes up the role of information about dispersion of beliefs for estimating bond market risk premiums. Finally, Section 8 presents concluding remarks.

2 Learning with Dynamic Term Structure Models

The standard learning environment within equilibrium, preference-based models of asset prices specifies agents' preferences and has them learning about the evolution of the state of the economy (e.g., Cogley and Sargent (2008) and Collin-Dufresne, Johannes, and Lochstoer (2013)). For several reasons, we depart from this literature and represent agents' learning problem in terms of a reduced-form pricing kernel. One reason is that estimated consumption-based models of the term structure often lead to large pricing errors,[7] whereas the volatilities

[6]Equally importantly, Dewachter and Lyrio (2008) assume constant market prices of risk in their Gaussian *DTSM*, which has the counterfactual implication that the expectations hypothesis holds. Also, the forecasts of yields from Laubach, Tetlow, and Williams (2007)'s *DTSM* are identical to those from a factor vector-autoregression (see JSZ), so the no-arbitrage structure of their *DTSM* is irrelevant for their learning problem.

[7]While there is recent progress on the development of nonlinear equilibrium models of the term structure, the evidence in Le, Singleton, and Dai (2010) and Joslin and Le (2013) suggests that equilibrium models within the affine class cannot successfully match the first and second conditional moments of bond yields.

of the pricing errors in reduced-form affine models are typically just a few basis points (Dai and Singleton (2000), Duffee (2002)). Equally importantly, bond-market participants are well known to use (often "affine") reduced-form models for pricing and risk assessment of their bond portfolios. Therefore, it seems at least as natural to explore learning within a framework that closely resembles those used by candidates for the marginal investor in bonds.

In our no-arbitrage setting we consider the set of portfolios of U.S. Treasury bonds with weights determined by the state-vector Z_t. In the absence of arbitrage opportunities and under weak regularity (closure) properties of the portfolio payoff space, there exists a unique stochastic discount factor (SDF) $\mathcal{M}(\Theta, \mathcal{P}_{t+1})$ that prices these Treasury portfolio payoffs (Hansen and Richard (1987)). The SDF $\mathcal{M}$ is governed by the parameter vector Θ and the pricing factors $\mathcal{P}_{t+1}$, and it implicitly depends on Z_t through the market prices of the risks $\mathcal{P}$. We assume that Z_t, which includes $\mathcal{P}_t$, follows a first-order Markov process. Under these assumptions, and absent the need to learn, the price D_t^m of a zero-coupon bond issued at date t and maturing at date m is the expected discounted payoff under the historical measure $\mathbb{P}$:

$$\begin{aligned} D_t^m &= E_t^{\mathbb{P}}\left[\prod\nolimits_{s=1}^{m}\mathcal{M}(\Theta, \mathcal{P}_{t+s})\right] \\ &= \int \prod\nolimits_{s=1}^{m}\mathcal{M}(\Theta, \mathcal{P}_{t+s})\, f_Z^{\mathbb{P}}(Z_{t+m}, \ldots, Z_{t+1}|Z_t; \Theta)\, dZ. \end{aligned} \tag{1}$$

We introduce learning into this arbitrage-free environment by working directly with the bond-market-specific SDF $\mathcal{M}(\Theta_t, Z_t)$, consistent with the vast majority of arbitrage-free *DTSMs* and the pricing strategies of many financial institutions. If instead one started with an equilibrium settings with learning about parameters (Cogley and Sargent (2008)), then an agent's SDF $\mathcal{K}(\gamma, Z_{t+1})$ for pricing nominal securities would typically be taken as given (γ is know to the agent) and this agent would be learning about the parameters governing the evolution of the state Z, $f_Z^{\mathbb{P}}(Z_{t+1}|Z_t, \beta)$. A bond-specific $\mathcal{M}(\Theta, \mathcal{P}_{t+1})$ implied by such a model is derived as $\mathcal{M}(\Theta, \mathcal{P}_{t+1}) = E[\mathcal{K}(\gamma, Z_{t+1})|\mathcal{P}_{t+1}, Z_t]$. Viewed this way, the Θ governing the reduced-form $\mathcal{M}$ is implicitly a confounding of γ and β. For this reason we assume that the representative agent who prices with $\mathcal{M}$ is learning about the entire parameter set Θ.

More generally, agents could be learning about Θ or Z_t. However, they certainly will have known that setting the risk factors in their pricing models to the low-order PCs of yields would allow them to capture the vast majority of the variation in bond yields, typically over 98% for bonds with m greater than one year and as few as three factors. Equally importantly, PCs are measured very accurately. Estimates of affine *DTSMs* (without learning) are virtually identical when the PCs are filtered from current yields or they are treated as observed without error (JSZ). Therefore, we focus on the problem of learning about Θ.

For our subsequent empirical analysis of learning we adopt a specific parametric form for $\mathcal{M}$ and law of motion for Z. Consider a bond market in which N yields are determined by three priced risks $\mathcal{P}_t$, with $3 < N$. We follow Joslin, Priebsch, and Singleton (2013) (JPS) and (without loss of generality) normalize $\mathcal{P}$ to be the first three PCs of bond yields, allow for an additional J variables M_t that influence the market prices of the risks $\mathcal{P}$, and assume that $Z_t' \equiv (\mathcal{P}_t', M_t')$ follows the Gaussian process

$$\begin{bmatrix} \mathcal{P}_{t+1} \\ M_{t+1} \end{bmatrix} = \begin{bmatrix} K_{\mathcal{P}0}^{\mathbb{P}} \\ K_{M0}^{\mathbb{P}} \end{bmatrix} + \begin{bmatrix} K_{\mathcal{P}\mathcal{P}}^{\mathbb{P}} & K_{\mathcal{P}M}^{\mathbb{P}} \\ K_{M\mathcal{P}}^{\mathbb{P}} & K_{MM}^{\mathbb{P}} \end{bmatrix} \begin{bmatrix} \mathcal{P}_t \\ M_t \end{bmatrix} + \Sigma_Z^{1/2} \begin{bmatrix} e_{\mathcal{P},t+1}^{\mathbb{P}} \\ e_{M,t+1}^{\mathbb{P}} \end{bmatrix}. \tag{2}$$

By the same reasoning underlying our assumption that Z_t is known to the consensus agent, we assume that the portfolio of yields $\mathcal{P}_t$ is priced perfectly by her pricing model. The relevant data for learning are then the history of $Z_1^t \equiv (Z_1, \ldots, Z_t)$ and of the $N-3$ linearly independent combinations, say $\mathcal{O}_1^t$, of yields that are not priced perfectly by $\mathcal{M}$.

The (logarithm of the) SDF $\mathcal{M}$ is assumed to take the form

$$\log \mathcal{M}_{t+1} = -r_t - \Lambda'_{\mathcal{P}t} e^{\mathbb{P}}_{\mathcal{P},t+1} - \frac{1}{2}\Lambda'_{\mathcal{P}t}\Lambda_{\mathcal{P}t}, \tag{3}$$

where r_t is the yield on a one-period bond. The market prices of $\mathcal{P}_t$ risks $\Lambda_{\mathcal{P}t}$,

$$\Lambda_{\mathcal{P}t} = \Lambda_0 + \Lambda_1 Z_t, \tag{4}$$

depend on Z_t (and not just $\mathcal{P}_t$) since agents will in general look beyond the bond market when assessing their willingness to bear risks indigenous to the bond market.

For future reference we note that, absent learning, the price of a zero-coupon bond with maturity m can be computed as the risk-neutral ($\mathbb{Q}$) expectation

$$D_t^m = E_t^{\mathbb{Q}}\left[\prod_{s=0}^{m-1} \exp(-r_{t+s})\right], \tag{5}$$

with the parameters normalized so that (see JSZ and Appendix B)

$$r_t = \rho_0 + \rho_{\mathcal{P}}\mathcal{P}_t, \tag{6}$$

and $\mathcal{P}_t$ follows the autonomous $\mathbb{Q}$-Gaussian process

$$\mathcal{P}_{t+1} = K^{\mathbb{Q}}_{0\mathcal{P}} + K^{\mathbb{Q}}_{\mathcal{P}\mathcal{P}}\mathcal{P}_t + \Sigma^{1/2}_{\mathcal{P}} e^{\mathbb{Q}}_{\mathcal{P},t+1}, \tag{7}$$

where ρ_0, $\rho_{\mathcal{P}}$, $K^{\mathbb{Q}}_{0\mathcal{P}}$, and $K^{\mathbb{Q}}_{\mathcal{P}\mathcal{P}}$ are all known functions $\left(r^{\mathbb{Q}}_{\infty}, \lambda^{\mathbb{Q}}, \Sigma_{\mathcal{P}}\right)$, with $r^{\mathbb{Q}}_{\infty}$ the risk-neutral mean of r_t, $\Sigma_{\mathcal{P}}$ the upper $K \times K$ block of Σ_Z, and $\lambda^{\mathbb{Q}}$ the K-vector of eigenvalues of $K^{\mathbb{Q}}_{\mathcal{P}\mathcal{P}}$. In this setting, bond yields take the affine form

$$y_t^m = A_m\left(r^{\mathbb{Q}}_{\infty}, \lambda^{\mathbb{Q}}, \Sigma_{\mathcal{P}}\right) + B_m\left(\lambda^{\mathbb{Q}}\right)\mathcal{P}_t. \tag{8}$$

For ease of notation we partition Θ into the two subsets of parameters: those governing the drift of Z in (2), $\Theta^{\mathbb{P}} \equiv (K_0^{\mathbb{P}}, K_Z^{\mathbb{P}})$; and those governing the pricing distribution, $\Theta^{\mathbb{Q}} \equiv (\lambda^{\mathbb{Q}}, r^{\mathbb{Q}}_{\infty}, \Sigma_Z)$, with the understanding that Σ_Z enters both the $\mathbb{P}$ and $\mathbb{Q}$ distributions.

The most general learning environment we consider captures real-time Bayesian learning and internally consistent pricing by the consensus agent.[8] Concretely, Bayesian learning has the following features:

[8]This is to be contrasted with bond pricing in the presence of time varying parameters as studied in the context of regime switching models (Ang and Bekaert (2002), Dai, Singleton, and Yang (2007)) with known parameters where bond prices reflect possible future changes in regimes. Our framework is also different from that of Feldhutter, Larsen, Munk, and Trolle (2012) who examine an optimal portfolio investment problem in the face of parameter uncertainty evaluated from full-sample estimation of a fixed-parameter affine *DTSM*.

- *Unknown Parameters*: the current value of Θ_t is unknown to the consensus agent, so she revises her view about Θ based on the distribution $f_\Theta^{\mathbb{P}}(\Theta_t^{t+m-1}|Z_1^t, \mathcal{O}_1^t)$.

- *Time Varying Parameters*: the unknown parameters may vary randomly over time.

- *No Compensation for Parameter Risk*: though Θ_t may be drifting and unknown, we assume that this risk is not priced in bond markets.

The absence of compensation for parameter risk is fairly standard in the literature on pricing with Bayesian learning, and it greatly simplifies what is already a challenging modeling problem. In particular, in our Bayesian learning models we take full account of learning about Θ_t when pricing bonds and forecasting excess returns– agents recognize explicitly that Θ_t is changing over time. This in contrast to the entire pricing literature that builds upon Cogley and Sargent (2008)'s notion of anticipated utility.

At each point in time the risk-neutral dynamics of $\mathcal{P}_t$ and r_t are assumed by the consensus agent to follow (6) and (7). The learning-augmented version of (5) is therefore

$$D_t^m = \int E^{\mathbb{Q}} \left[\prod_{s=1}^m exp(-r_{t+s}) | \Theta_t^{\mathbb{Q},t+m-1} \right] f_\Theta^{\mathbb{Q}} \left(\Theta_t^{\mathbb{Q},t+m-1} | Z_1^t, \mathcal{O}_1^t \right). \tag{9}$$

Bond prices under learning are not described by an affine *DTSM* even when Z follows an affine process. Nevertheless, we show in Appendix E that, conditional on a deterministic path of the risk neutral coefficients $\Theta_t^{\mathbb{Q},t+m-1}$, bond yields will still be affine in $\mathcal{P}_t$. This fact allows us to compute model-implied bond yields by solving the conditionally affine models for each path of $\Theta_t^{\mathbb{Q},t+m-1}$, and then integrating over the distribution of possible future $\mathbb{Q}$ paths.

To complete the asset pricing framework we assume that, given the agent's views $\Theta_t^{\mathbb{P}}$ and Σ_{Zt} at date t, the conditional $\mathbb{P}$-distribution of Z_{t+1} is given by

$$f_Z^{\mathbb{P}}(Z_{t+1}|Z_1^t) = N\left(K_{0t}^{\mathbb{P}} + K_{Zt}^{\mathbb{P}} Z_t, \Sigma_{Zt} \right); \tag{10}$$

and we define the SDF as

$$\log \mathcal{M}_{t+1} = -r_t - \Lambda_{\mathcal{P}t} e_{\mathcal{P},t+1}^{\mathbb{P}} - \frac{1}{2} \Lambda'_{\mathcal{P}t} \Lambda_{\mathcal{P}t} \tag{11}$$

$$\Lambda_{\mathcal{P}t} = \Lambda_{0t}(\Theta_t) + \Lambda_{1t}(\Theta_t) Z_t, \tag{12}$$

where $\Lambda_{\mathcal{P}t}$ depends on the entire vector Θ. Under the assumption of constant and known parameters, this learning model simplifies to the standard Gaussian *DTSM*.

At date t the Bayesian agent, faced with new observations $(Z_t, \mathcal{O}_t)$ and the past history $(Z_1^{t-1}, \mathcal{O}_1^{t-1})$, evaluates the (approximate) likelihood function by updating her posterior view on the unobserved parameters and then integrating over this posterior distribution:

$$\begin{aligned} f(Z_1^t, O_1^t) \;\; = \;\; & \prod_{s=1}^t \int f(Z_s | Z_1^{s-1}, \mathcal{O}_1^{s-1}, \Theta_{s-1}^{\mathbb{P}}, \Sigma_{Z,s-1}) f(\mathcal{O}_s | Z_1^s, O_1^{s-1}, \Theta_s^{\mathbb{Q},s+m-1}, \Sigma_{\mathcal{O}}) \\ & f(\Theta_{s-1}^{\mathbb{P}}, \Sigma_{Z,s-1}, \Theta_s^{\mathbb{Q},s+m-1}, \Sigma_{\mathcal{O}} | Z_1^{s-1}, \mathcal{O}_1^{s-1}) d(\Theta_{s-1}^{\mathbb{P}}, \Sigma_{s-1}, \Theta_s^{\mathbb{Q},s+m-1}, \Sigma_{\mathcal{O}}), \end{aligned} \tag{13}$$

where $\Sigma_{\mathcal{O}}$ is the variance of the pricing errors on $\mathcal{O}_t$ which agents view as fixed but unknown. Regarding the form of the posterior density of the parameters, $(\Theta^{\mathbb{P}}_{s-1}, \Sigma_{Z,s-1})$ govern the $\mathbb{P}$-distribution of Z and the vector $\Theta^{\mathbb{Q},s+m-1}_s$ governs the pricing of the yield portfolios $\mathcal{O}_s$. The specific form of this density depends on the nature of the learning problem. For instance, if the consensus agent views $r^{\mathbb{Q}}_\infty$ and $\Sigma_{\mathcal{PP}}$ as fixed (though still unknown) in the underlying economy, then $\Theta^{\mathbb{Q},t+m-1}_t$ is short-hand notation for the triple $(\lambda^{\mathbb{Q},t+m-1}_t, r^{\mathbb{Q}}_t, \Sigma_{\mathcal{P}t})$, as only her views about $(r^{\mathbb{Q}}, \Sigma_{\mathcal{P}})$ as of date t affect pricing.

2.1 Naive *DTSM*-Based Learning Rules

Relatively naive/myopic learning rules are easily derived from (13) by specializing the posterior density of $(\Theta^{\mathbb{P}}_{s-1}, \Sigma_{Z,s-1}, \Theta^{\mathbb{Q},s+m-1}_s, \Sigma_{\mathcal{O}})$. We focus on two that we believe are practically relevant and that, as we will see, turn out to match well with the learning behavior of professionals. The first, which we will refer to simply as the *naive* rule, has the consensus agent updating $(\Theta, \Sigma_{\mathcal{O}})$ every month with the ML estimator $(\widehat{\Theta}_t, \widehat{\Sigma}_{\mathcal{O}})$ from a standard fixed-parameter Gaussian $DTSM$ using the then available history $(Z^t_1, \mathcal{O}^t_1)$. That is, every month $\widehat{\Theta}_t$ is chosen to maximize

$$f(Z^t_1, O^t_1; \Theta) \;\; = \;\; \prod_{s=1}^{t} f(Z_s|Z^{s-1}_1, \mathcal{O}^{s-1}_1; \Theta^{\mathbb{P}}, \Sigma_Z) f(\mathcal{O}_s|Z^s_1, O^{s-1}_1; \Theta^{\mathbb{Q}}, \Sigma_{\mathcal{O}}). \tag{14}$$

What this strategy amounts to is completely ignoring the uncertainty about Θ when both updating one's views about Θ and in pricing. As such it is naive for both forward- and backward-looking reasons. Looking forward, bond prices implicitly depend on future $\mathbb{Q}$-forecasts through the standard affine pricing formulas (8) evaluated at $\widehat{\Theta}_t$, and the dependence of these forecasts on future $\Theta^{\mathbb{Q}}$ is addressed by fixing future $\Theta^{\mathbb{Q}}_s$, $s > t$, at $\widehat{\Theta}^{\mathbb{Q}}_t$:

$$f^{\mathbb{Q}}\left(\Theta^{\mathbb{Q},t+m-1}_t | Z^t_1, \mathcal{O}^t_1\right) \equiv f^{\mathbb{Q}}\left(\Theta^{\mathbb{Q}}_{t+m-1} = \widehat{\Theta}^{\mathbb{Q}}_t, \ldots, \Theta^{\mathbb{Q}}_{t+1} = \widehat{\Theta}^{\mathbb{Q}}_t, \Theta^{\mathbb{Q}}_t = \widehat{\Theta}^{\mathbb{Q}}_t | Z^t_1, \mathcal{O}^t_1\right). \tag{15}$$

Similarly, $\mathbb{P}$-forecasts of future bond yields, and hence risk premiums, are based on the fitted vector-autoregression (VAR) (2) assuming that Θ is fixed at the current estimate $\widehat{\Theta}_t$ *even though* $\widehat{\Theta}_{t+1}$ *will in fact change with the arrival of new information.* This learning rule is also naive looking backwards, because $\widehat{\Theta}_t$ is updated by estimating a likelihood function over the sample up to date t presuming that Θ is fixed and *has never changed in the past even though* $\widehat{\Theta}_t$ *did change every month.*

There is a revealing middle ground between this naive learning rule and that of the Bayesian agent, one in which the learner is "semi-consistent" ($\mathcal{SC}$) in updating and pricing. Starting again from (13), we suppose that agent treats $(\Theta^{\mathbb{Q}}, \Sigma_{\mathcal{O}})$ as known in updating and pricing, but recognizes the need to learn about $\Theta^{\mathbb{P}}$ for the purpose of both pricing and forecasting. For this agent the relevant likelihood function of the data becomes:

$$f(Z^t_1, O^t_1) \;\; = \;\; \prod_{s=1}^{t} f(\mathcal{O}_s|Z^s_1, O^{s-1}_1; \Theta^{\mathbb{Q}}, \Sigma_{\mathcal{O}}) \times \int f(Z_s|Z^{s-1}_1, \mathcal{O}^{s-1}_1, \Theta^{\mathbb{P}}_{s-1}; \Sigma_Z) f(\Theta^{\mathbb{P}}_{s-1}|Z^{s-1}_1, \mathcal{O}^{s-1}_1) d(\Theta^{\mathbb{P}}_{s-1}). \tag{16}$$

This intermediate case is interesting for three reasons: (i) its structure can be reinterpreted as a constrained case of the fully Bayesian rule; (ii) the presumption that $\Theta^{\mathbb{Q}}$ is fixed will turn out to be close to being satisfied by our empirical learning rules; and (iii) because of this feature of $\widehat{\Theta}^{\mathbb{Q}}$, the rules for the (naive) ML-based and the (constrained Bayesian) semi-consistent learners will turn out to be nearly identical empirically.

Elaborating for a concrete case, consider a $\mathcal{SC}$ agent who perceives herself as learning about $\Theta^{\mathbb{P}}$ taking $(\Theta^{\mathbb{Q}}, \Sigma_{\mathcal{O}})$ as known. Suppose that $\Theta_t^{\mathbb{P}}$ can be partitioned as $(\psi^r, \psi_t^{\mathbb{P}})$, where $\psi_t^{\mathbb{P}}$ is the vectorized set of free parameters and ψ^r is the vectorized set of parameters that are fixed conditional on $\Theta^{\mathbb{Q}}$. The constraints we impose subsequently on the market prices of $\mathcal{P}$-risks admit this partition. Then (10) becomes (see Appendix C for details)

$$f_y^{\mathbb{P}}(y_{t+1}|Z_1^t) = N\left(x_t\psi_t^{\mathbb{P}}, \Sigma_y\right), \tag{17}$$

where $y_t = Z_t - \left(I \otimes [1, Z_{t-1}']\right)\iota_r\psi^r$, and $x_t = (I \otimes [1, Z_t'])\,\iota_f$, with ι_r and ι_f denoting the matrices that select the columns of $(I \otimes [1, Z_{t-1}'])$ corresponding to the restricted and free parameters, respectively. Let $\widehat{\psi}_t^{\mathbb{P}}$ denote an estimator of $\psi_t^{\mathbb{P}}$ in (17), viewed as a subvector of the estimator $\widehat{\Theta}_t^{\mathbb{P}}$.

For the purpose of comparing the $\mathcal{SC}$ and naive learning rules we focus on the case where the agent believes that $\psi_t^{\mathbb{P}}$ evolves according to the process

$$\psi_t^{\mathbb{P}} = \psi_{t-1}^{\mathbb{P}} + \eta_t, \qquad \eta_t \overset{iid}{\sim} N(0, Q_t), \tag{18}$$

where Q_t denotes the (possibly) time-varying covariance matrix of η_t. We assume that η_t is independent of all past and future $e_{Zt}^{\mathbb{P}}$. The construction of the semi-consistent learning rule involves filtering on $\psi^{\mathbb{P}}$ conditional on $(\Theta^{\mathbb{Q}}, \psi^r)$. Adopting a Gaussian prior on $\psi_0^{\mathbb{P}}$ leads to a posterior distribution for $\psi_t^{\mathbb{P}}$ that is also Gaussian, $\psi_t^{\mathbb{P}}|Z_1^t \sim N(\mu_t, P_t)$, with P_t is defined in Appendix C and the posterior mean following the recursion

$$\mu_t = \mu_{t-1} + R_t^{-1}x_{t-1}'\Sigma_Z^{-1}(y_t - x_{t-1}\mu_{t-1}), \tag{19}$$

where $R_t^{-1} \equiv P_t - Q_t$ and R_t satisfies the recursion (see Appendix C)

$$R_t = \left(I - P_{t-1}^{-1}Q_{t-1}\right)R_{t-1} + x_{t-1}'\Sigma_Z^{-1}x_{t-1}. \tag{20}$$

Notably, this $\mathcal{SC}$ rule can interpreted within the class of *adaptive least-squares estimators* (ALS) of $\psi_t^{\mathbb{P}}$. We say that $\widehat{\psi}_t^{\mathbb{P}}$ is an ALS estimator if there exists a sequence of scalars $\gamma_t > 0$ such that $\widehat{\psi}_t^{\mathbb{P}}$ can be expressed recursively as

$$\widehat{\psi}_t^{\mathbb{P}} = \widehat{\psi}_{t-1}^{\mathbb{P}} + R_t^{-1}x_{t-1}'\Sigma_Z^{-1}(y_t - x_{t-1}\widehat{\psi}_{t-1}^{\mathbb{P}}), \tag{21}$$

$$R_t = \gamma_t R_{t-1} + x_{t-1}'\Sigma_Z^{-1}x_{t-1}. \tag{22}$$

It follows immediately from (19) - (20) that the posterior mean in the Kalman filter used by the Bayesian learner to update $\psi_t^{\mathbb{P}}$ can be represented as a generalized ALS estimator. Moreover, (20) reveals three special cases where the filtering underlying $\mathcal{SC}$ learning reduces to an ALS estimator (that is, (20) reduces to (22)):[9]

$\mathcal{B}{\downarrow}$**ALS:** Setting $P_{t-1}^{-1}Q_{t-1} = (1-\delta_t)\cdot I$[10] for some sequence of scalars $0 < \delta_t \le 1$, μ_t becomes

[9]See McCulloch (2007), and the references therein, for discussions of similar issues in a setting of univariate y_t and econometrically exogenous x_t.

[10]This condition can be obtained by recursively setting $Q_t = (\delta_{t+1}^{-1} - 1)(P_{t-1} - P_{t-1}x_{t-1}'\Omega_t^{-1}x_{t-1}P_{t-1})$.

an *ALS* estimator of $\psi^{\mathbb{P}}$ with $\gamma_t = \delta_t$.

$\mathcal{B}\downarrow$**CGLS:** Specializing further by setting $\delta_t = \delta$ to a constant leads to μ_t being a *constant gain least-squares (CGLS)* estimator of $\psi^{\mathbb{P}}$ with $\gamma = \delta$.

$\mathcal{B}\downarrow$**RLS:** If $\delta_t = \gamma = 1$, then we recover the *recursive least-squares (RLS)* estimator of $\psi^{\mathbb{P}}$.

Among the insights that emerge from this construction is that an $\mathcal{SC}$ agent whose learning rule specializes to the *RLS* estimator is not adaptive in the following potentially important sense. With $\gamma = 1$ it follows that $Q_t = 0$, so an agent exhibiting *RLS* updating on $\psi^{\mathbb{P}}$ presumes that these parameters are fixed over time. Learning is only about the unknown fixed value of $\psi^{\mathbb{P}}$. Consequently, sudden changes in market conditions that result in sharp movements in recent values of Z may have an imperceptible effect on $\widehat{\psi}_t^{\mathbb{P}}$ as updated by this agent. Indeed, in environments where the ML estimator converges to a constant for large T, as is typically presumed by regularity to ensure consistency of ML estimation in non-learning environments, then after a very long training period we would expect this agent to be virtually non-adaptive on $\widehat{\psi}^{\mathbb{P}}$ to new information.

An $\mathcal{SC}$ rule that is more accommodating to changes in the underlying structure of the economy (say to changes in monetary or other policies in response to changing business conditions or in the face of financial crises) is obtained by giving less weight to values of Z far in the past. Such down-weighting arises naturally when the $\mathcal{SC}$ learner specializes to Case $\mathcal{B}\downarrow$**CGLS** of *CGLS* learning. The constant-gain coefficient γ determines the “half-life” of the down weighting of the past history. While this might not be immediately apparent from the likelihood function implied by $\mathcal{SC}$ learning with *ALS* updating in Appendix C, it can be seen heuristically by noting that, conditional on $\Theta^{\mathbb{Q}}$, the first-order conditions to this likelihood function are identical to those of the scaled Gaussian likelihood of the naive learner in which the quadratic form in $u_t \equiv y_t - x_{t-1}\psi_{t-1}^{\mathbb{P}}$ is weighted by γ^t; that is, the likelihood involves terms of the form $\gamma^t u_t' \widehat{\Sigma}_y^{-1} u_t$.

Summarizing these observations, if there are no constraints on the market prices of risk $\Lambda_{\mathcal{P}}$ (ψ^r is empty) then, in each period, ML estimates of $\Theta^{\mathbb{P}}$ used by the naive learner are identical to the ordinary least-squares estimator from the VAR for Z_t (JPS). What the preceding derivations show is that the naive learner behaves exactly as a $\mathcal{SC}$ learner under the presumption that $\Theta^{\mathbb{P}}$ is fixed (but unknown). Put differently, the *RLS* updating of $\Theta^{\mathbb{P}}$ by the naive learner is actually exactly how a Bayesian learner would update under the presumption of known $(\Theta^{\mathbb{Q}}, \Sigma_Z, \Sigma_{\mathcal{O}})$ and under the constraints on the posterior parameter distribution that give rise to Case $\mathcal{B}\downarrow$**RLS**. Similarly, with ψ^r empty and under Case $\mathcal{B}\downarrow$**CGLS**, updating by *CGLS* is exactly how a Bayesian with full knowledge of $(\Theta^{\mathbb{Q}}, \Sigma_Z, \Sigma_{\mathcal{O}})$ would update her views on $\Theta_t^{\mathbb{P}}$.

Subtle differences across rules arise when there are constraints on $\Lambda_{\mathcal{P}}$ (ψ^r is non-empty). Most sophisticated is the Bayesian agent who recognizes that Θ is unknown and may be changing over time and builds this knowledge into learning, pricing, and forecasting. Less sophisticated is the $\mathcal{SC}$ agent who estimates $\Theta_t^{\mathbb{Q}}$ say by ML, but then treats the resulting $\widehat{\Theta}_t^{\mathbb{Q}}$ as known and fixed when learning about (updating) $\widehat{\Theta}_t^{\mathbb{P}}$ according to one of the schemes nested in Case $\mathcal{B}\downarrow$**CGLS**. Least sophisticated of all, in principle, is the agent who updates $\widehat{\Theta}_t$ by re-estimating the *DTSM* by ML every month as new data becomes available, either directly or (analogously to Case $\mathcal{B}{\downarrow}\mathbf{CGLS}$) by down weighting the quadratic forms in the likelihood function by γ^t as described above. Even though naive and $\mathcal{SC}$ learning schemes give similar updates on $\Theta^{\mathbb{P}}$, they are not identical because they imply different dependencies among $\widehat{\psi}^{\mathbb{P}}$ and $\widehat{\Theta}^{\mathbb{Q}}$. Henceforth we focus on the $\mathcal{SC}$ formulation of the likelihood function, that emerges as a constrained special case of the Bayesian learner's estimation problem. In practice, we obtain nearly identical fitted learning rules from implementing their naive counterparts.

2.2 Learning and Heterogeneity of Beliefs

Up to this point we have framed our discussion of learning in terms of the rule followed by a consensus agent, unlike in most equilibrium models of bond prices (e.g., Xiong and Yan (2009) and Buraschi and Whelan (2012)). While our learning environment is quite rich, we have abstracted completely from the potentially rich heterogeneity across market participants. In fact, there is considerable cross-sectional dispersion of the time-series of forecasts of professionals in bond markets. Conceptually, the SDF $\mathcal{M}$ exists even in the presence of such heterogeneity as an implication of the absence of arbitrage opportunities. However we are adding the (strong) additional structure that, in the presence of heterogeneity, bond prices are representable in terms of the reduced-form SDF of an investor who holds the consensus beliefs of market professionals.

While, in general, the learning problems faced by market participants need not aggregate up to a representation in terms of a consensus agent's learning problem as in (9), when markets are complete there typically exists a fictitious "aggregate agent" who holds consensus beliefs and whose SDF prices the term structure of yields (Jouini and Napp (2007)). This result is key to the tractability of prior equilibrium models with heterogeneous investors. However, a major concern with this rationalization of our learning rules is that aggregation leads to an SDF for the consensus agent that depends on a priced "aggregation bias" term related to the differences in beliefs across the market participants. Typically (see, e.g., Buraschi and Whelan (2012)) this bias term depends on the pairwise differences in agents' beliefs and, thereby, indirectly introduces a very high-dimensional set of priced risk factors. This high dimensionality seems counterfactual in the light of the enormous empirical success of low-dimensional factor models for bond yields.

This last point is material as including measures of forecaster disagreement directly as risk factors in $\mathcal{P}$ would likely lead to over-parameterized models with detrimental consequences for goodness-of-fit. *DTSMs* that keep $dim(\mathcal{P})$ small and introduce macro variables as risk factors typically do not accurately price bonds, whereas *DTSMs* with *PC*s as the risk factors have very small pricing errors. Furthermore, Duffee (2010) shows that the over-fitting associated with a large number of risk factors (four or five in his Gaussian *DTSMs*) results in model-implied Sharpe ratios for bond portfolios that are implausibly large.

Moreover, our learning environment is agnostic as to the underlying economic sources of variation in $\mathcal{P}_t$. The *PC*s may already encompass the relevant (priced in bond markets) aggregation bias that arises in equilibrium models. For a rough assessment of the extent to which heterogeneity in beliefs is spanned by $\mathcal{P}$, we constructed measures of the cross-sectional dispersion in professionals' forecasts of $\mathcal{P}$, $MAD(PCi)$, for $i = 1, 2, 3$. At each date t, and for a given forecast horizon, we determined the cross-sectional median forecast of $\mathcal{P}$, and then

	$\mathcal{P}$				$(\mathcal{P}, M)$			
	1Q	2Q	3Q	4Q	1Q	2Q	3Q	4Q
$MAD(PC1)$	0.32	0.50	0.58	0.58	0.34	0.50	0.57	0.58
$MAD(PC2)$	0.54	0.61	0.62	0.61	0.66	0.67	0.68	0.69
$MAD(PC3)$	0.65	0.66	0.68	0.69	0.66	0.67	0.69	0.69
$MAD(INF)$	0.04	0.04	0.13	0.27	0.40	0.30	0.28	0.32

Table 1: Adjusted R^2's from the projections of the cross-forecaster mean absolute errors in forecasting the PCs onto $\mathcal{P}$ and onto $(\mathcal{P}, M)$ over the forecast horizons of one through four quarters. The sample period is January, 1985 through March, 2012.

computed $MAD(PCi)$ as the cross-sectional mean absolute deviations of individual forecasts from the median. These dispersion measures $MADF(PCi)$ were then projected onto the contemporaneous values $\mathcal{P}_t$ and the broader state vector $(\mathcal{P}_t, M_t)$. An analogous calculation, $MAD(INF)$, was done using the cross-section of inflation forecasts.

From Table 1 it is seen that the levels of the PC's and the $MAD(PCi)$ are highly correlated, with $\mathcal{P}$ accounting for over half of the variation in the dispersion measures for forecast horizons beyond two quarters (the only exception is the one-quarter ahead forecast of $PC1$). Moreover, the percentages of variation in the MAD's explained by $\mathcal{P}$ increase with the order of the PC, with nearly seventy percent of the variation in the professional disagreement about the future path of the curvature factor ($PC3$) being spanned by $\mathcal{P}_t$. Interestingly, from the right side of Table 1 one sees that adding in the conditioning information (INF, NAI) adds little beyond the explanatory power of $\mathcal{P}$ for $MAD(PC1)$ or $MAD(PC3)$. In contrast, M_t has incremental explanatory power for $MAD(PC2)$, the dispersion in professional views about the slope of the yield curve.

Ultimately, for addressing the issue of heterogeneity, we are interested in whether measures such as $MAD(PCi)$ embody incremental information beyond Z_t about risk premiums in bond markets. We defer this issue until Section 7 in part because the survey forecast data is only available for the later portion of our sample. We prefer to explore learning with a generous training period for initializing the learning rules. Additionally, we intentionally focus on learning rules that do not condition on survey information with the goal of providing a characterization of how the historical consensus survey forecasts are formed.

3 Learning Rules, Empirical Constructs, Historical Data

As references for exploring the properties of learning rules in bond markets we use the survey forecasts of yields and inflation by market professionals from the Blue Chip Financial Forecasters (BCFF) over the period January, 1985 through March, 2012. This survey is typically released at the beginning of the month (usually the first business day), based on information collected over a two-day period (usually between the 20th and the 26th of the previous month). A total of 177 institutions provide forecasts during the period of our study. These institutions are divided by the BCFF survey provider into different broad sectors: Financial Services, Consulting, Business Associations, Manufacturing, Insurance

Companies and Universities. In our sample there are 48 consulting companies, 111 financial service companies, and the remaining 18 are from insurance companies, business associations, manufacturers, and universities. Forecasts are averages over calendar quarters and cover horizons out to five quarters ahead. For example, in January, 1999, the one-quarter ahead forecast for a specific variable will be equal to its average value over February and March; the two-quarters ahead forecast will be for the average value between April and June, and so on.

The survey forecasts are for U.S. Treasury 6-month bill yield and par yields on coupon bonds with maturities of 1, 2, 3, 5, 7, and 10 years. One of our ultimate goals is to compare the forecasts that emerge from our *DTSMs* with learning to the consensus forecasts of these professionals. (We do not use the survey data directly in the estimation of our learning models.) Accordingly, we use the survey-implied forecasts of averages of zero-coupon bond yields computed by Le and Singleton (2012). This allows for direct comparisons of the zero yield forecasts implied by a *DTSM* to their counterparts from survey data.[11]

For our empirical implementation, yields on zero-coupon bonds with maturities of 6 months and 1, 2, 3, 5, 7, and 10 years were calculated from coupon-bond yields as reported in the CRSP database using the Fama-Bliss methodology for the sample period June, 1961 through March, 2012.[12] For M, inflation is measured as the twelve-month moving average of monthly CPI core inflation (INF), and real growth is measured by the three-month moving average of a real activity index (NAI) that we construct in manner similar to the National Activity Index reported by the Federal Reserve Bank of Chicago. Specifically, we use the first PC of the monthly unemployment, industrial production, consumption of durable and non-durable goods, and non-farm workers hourly income, obtained from the Bureau of Labour Statistics.[13]

Three classes of learning rules are examined. The first, denoted by $\ell(RW)$, constructs forecasts of future zero yields according to the random walk model which conditions only on the lagged value of the yield being forecasted. The second, $\mathcal{SC}$ and *DTSM*-based rule, denoted $\ell(JSZ)$, is based on the three-factor model developed in JSZ in which $\mathcal{P}$ is comprised of the first three PC's of the zero yields, measured without error, and the market prices $\Lambda_{\mathcal{P}}$ of these risks are affine functions of $\mathcal{P}_t$. The third rule, denoted $\ell(JPS)$, is based on the model in JPS, which extends JSZ by allowing for macroeconomic information M to influence $\Lambda_{\mathcal{P}}$, and in our setting M includes measures of inflation and real output growth (see below). These *DTSM*-based rules enforce the constraint the market price of $PC3$ is zero, as in the preferred model of JPS.[14] Since $K_Z^{\mathbb{Q}}$ is fully determined by $\lambda^{\mathbb{Q}}$, constraints on $\Lambda_{\mathcal{P}}$ effectively transfer *a priori* knowledge of $\lambda^{\mathbb{Q}}$ to (some) knowledge about $K_Z^{\mathbb{P}}$.

These learning rules are initialized using estimates obtained for the sample period June, 1961 through June 1975. Then every month, up through March, 2011, as new data becomes

[11]Whereas forecasting zero-coupon yields in an affine *DTSM* is a linear forecasting problem (see below), par yields are nonlinear functions of zero-coupon yields. We avoid this complexity by interpolating the forecasts of par yields to obtain approximate forecasts of zero yields.

[12]See Le and Singleton (2012) for details and the reasons they constructed this data from the original Treasury bond yields rather than use other widely available splined zero-coupon yields.

[13]We do not use the index from the Chicago Federal Reserve, because it does not extend as far back in time as our yield data. For the overlapping sample period of March, 1967 through March, 2012, the correlation of our index and the CFNAI is 0.8.

[14]JPS also impose a constraint on the eigenvalues of $K_1^{\mathbb{Q}}$ and $K_1^{\mathbb{P}}$. The potential role of this constraint in our learning rules is taken up in Section 4.

available, the agent re-estimates the model and bases forecasts on the current month's estimates. Moving through this period, we envision the consensus agent implementing the JSZ normalizations for the *DTSM*-based rules every month based on current information. Thus, the portfolio weights for the first three *PC*s are recomputed each month from the estimated variance-covariance matrix of yields.[15]

From the fitted *DTSM* at date t, an h-period ahead forecast of Z is given by

$$\hat{Z}_{t+h} = \widehat{K}_{0t}^{\mathbb{P}} + \left(\widehat{K}_{Zt}^{\mathbb{P}}\right) \widehat{K}_{0t}^{\mathbb{P}} + ... + \left(\widehat{K}_{Zt}^{\mathbb{P}}\right)^{h-1} \widehat{K}_{0t}^{\mathbb{P}} + \left(\widehat{K}_{Zt}^{\mathbb{P}}\right)^{h} Z_t. \tag{23}$$

This leads directly to the h-period ahead forecasts of yields:

$$\hat{y}_{t+h}^{m} = A_m\left(\widehat{K}_{0t}^{\mathbb{Q}}, \widehat{K}_{Zt}^{\mathbb{Q}}, \widehat{\Sigma}_{\mathcal{P}\mathcal{P}t}\right) + B_m\left(\widehat{K}_{Zt}^{\mathbb{Q}}\right)\hat{\mathcal{P}}_{t+h}. \tag{24}$$

If h is the last month in a quarter, then the average expected yield over the quarter is:

$$\hat{y}_{t+h:t+h-3}^{m} = \frac{1}{3}\sum_{i=1}^{3} \hat{y}_{t+h-i}^{m}. \tag{25}$$

We compare these forecasts to the median of the forecasts reported by BCFF professionals.

One of our objectives is to characterize the implicit learning rule $\ell(BCFF)$ in terms of a *DTSM*-based rule. Precisely how professionals forecast future yields is unknown, though it is reasonable to presume that they condition on more information than what is encompassed by Z. We do know that the forecasts that emerge from our *DTSM*-based learning rules are inefficient in the sense that they ignore the forecaster's knowledge that $\widehat{\Theta}_t$ is being revised every period. Comparisons are further challenged by the fact that the various learning rules often differ in their forecasting accuracy depending on what aspect of yield-curve shape or forecast horizon one is focused on. With this in mind we present results for various learning rules, for both Cases $\mathcal{B}$↓**RLS** and $\mathcal{B}$↓**CGLS**. As noted previously, the latter is more adaptable to structural change or parameter drift. A summary of the learning rules that are implemented empirically is displayed in Table 2.

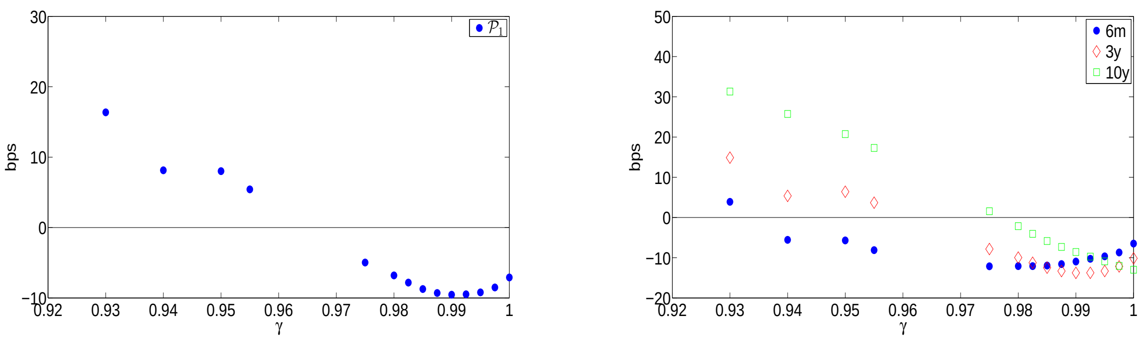
To replicate the implicit rule followed by BCFF professionals, we considered an $\mathcal{SC}$ agent satisfying Case $\mathcal{B}$↓**CGLS** and searched for the value of γ that gave the best match of $\ell(JSZ_{CG})$ to $\ell(BCFF)$. Specifically, we examined γ's in the range of $[0.94, 1.00]$ and, for each γ, we computed the RMSE's of the differences in the errors in forecasting $PC1$ and individual bond yields one year ahead from the BCFF- and JSZ-based rules. We found, both for the individual bonds and $PC1$, that $\gamma = 1$ (*RLS* updating) produced the best replicating rule. This was true for the entire sample between 1985 and 2011, and also for the subsamples of 1985-1999 and 2000-2011. The replicating rule is denoted by $\ell(JSZ)$ in Table 2.

The finding that $\gamma = 1$ gives the best tracking rule for $\ell(BCFF)$ raises the question of whether there is a rule $\ell(JSZ_{CG})$ with $\gamma < 1$ that systematically outperforms rule $\ell(BCFF)$ in terms of RMSE's of forecasts. To assess this we examined the differences between the RMSE's for $PC1$ and individual yields and searched for the value of γ that gave the best

[15] This is in contrast to the more typical practice of of using full-sample estimates of the loadings for the *PC*s when estimating and forecasting with *DTSMs*.

Rule	*DTSM*	Information	Restrictions	γ
$\ell(RW)$	Random Walk	Own Yield	N/A	N/A
$\ell(JSZ)$	Joslin, Singleton, and Zhu (2011)	$\mathcal{P}$	No-Arbitrage $PC3$ unpriced	1
$\ell(JSZ_{CG})$	Joslin, Singleton, and Zhu (2011)	$\mathcal{P}$	No-Arbitrage + $PC3$ unpriced	0.99
$\ell(JPS)$	Joslin, Priebsch, and Singleton (2013)	$(\mathcal{P}, M)$	No-Arbitrage + $PC3$ unpriced	1

Table 2: Description of models used as learning rules.



(a) Forecasting $PC1$ (b) Forecasting Individual Yields

Figure 1: Comparison of differences in RMSE's for (a) forecasting $PC1$ and (b) individual yields over a one-year horizon between the rules $\ell(JSZ_{CG})$ and $\ell(BCFF)$. The sample is January, 1985 through March, 2012.

relative performance of (largest negative difference between the RMSE's of) $\ell(JSZ_{CG})$ over $\ell(BCFF)$). Figure 1 shows that, for $PC1$ and the yields on the bond with a maturity of three years, the optimal choice of γ for $CGLS$ learning is 0.99. Interestingly, the optimal γ for the six-month yield was much smaller, roughly 0.97, while the optimal γ for the ten-year yield was again 1.00. We infer from these results that optimal learning, at least within the family $\ell(JSZ_{CG})$, varies by the maturity of the bond with a consensus $\mathcal{SC}$ learner preferring to accommodate more drift in the fitted rules the shorter is the maturity of the bond, holding fixed the forecast horizon of one year. Henceforth, $\ell(JSZ_{CG})$ refers to the rule with $\gamma = 0.99$.

4 Naive Learning Rules

We begin our empirical exploration of learning by assuming that the $\mathcal{SC}$ consensus agent uses the naive *DTSM*-based rules introduced in Section 2.1 and summarized in Table 2. We know from the large literature on Gaussian *DTSMs* without learning that in-sample pricing is very accurate. In our learning environment agents have the added flexibility of updating $\widehat{\Theta}_t$ every month. If updates of $\widehat{\Theta}_t^{\mathbb{Q}}$ are large, then the naive learning rules could show notably

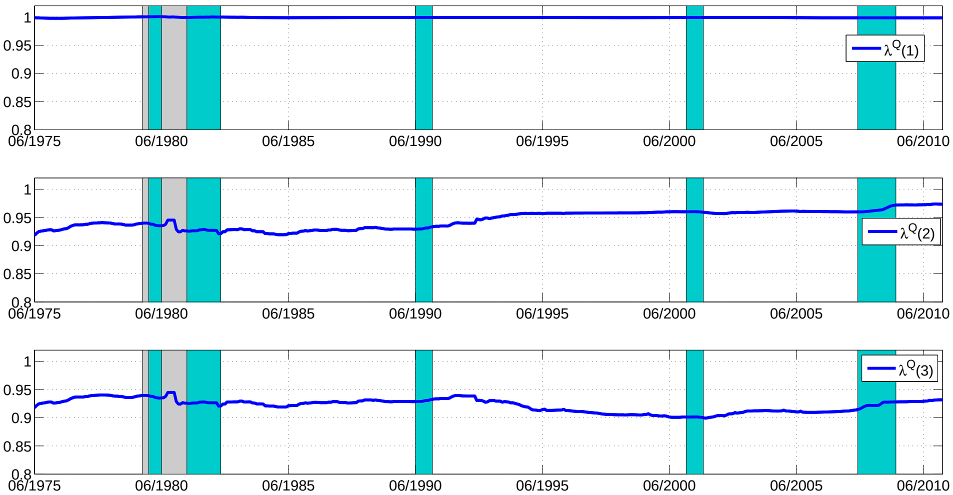


Figure 2: Estimates from model $\ell(JSZ)$ of the eigenvalues $\lambda^{\mathbb{Q}}$ of $K_1^{\mathbb{Q}}$ that govern the loadings on $\mathcal{P}_t$ in the affine representations of bond yields. The estimates at date t are based on the historical data up to observation t, over the period June, 1975 to March, 2012.

inferior pricing accuracy relative to professionals, who implicitly use richer information sets, and Bayesian learners who formally accommodate the uncertain future paths of $\widehat{\Theta}_t$ when computing prices.

Delving deeper into pricing accuracy, note that agents can reasonably be modeled as knowing $\lambda^{\mathbb{Q}}$. This is because the B_m, which depend only on $\lambda^{\mathbb{Q}}$, are pinned down very precisely by the cross-section of yields;[16] the historical time series information is largely irrelevant for learning about $\lambda^{\mathbb{Q}}$. Not only is it reasonable to presume that $\lambda^{\mathbb{Q}}$ is largely known, it turns out that revising $\widehat{\Theta}_t$ according to a *DTSM*-based rule leads one to hold $\widehat{\lambda}_t^{\mathbb{Q}}$ essentially fixed over time. This can be seen from Figure 2 for the rule $\ell(JSZ)$.[17] Indeed, repeating our learning exercise with $\lambda^{\mathbb{Q}}$ fixed from the initial training period onward has virtually no effect on the properties of the rule-implied prices or forecasts.

Pursuing this insight, if $\lambda^{\mathbb{Q}}$ is known and fixed over time, then so are the loadings B_m on $\mathcal{P}$ in the affine pricing expression (8). Combining this with the fact that $\mathcal{P}_t$ is measured with negligible error, the state-dependent components of bond yields that emerge from (8) with learning take the same form $B_m\left(\lambda^{\mathbb{Q}}\right)\mathcal{P}_t$, just as in a *DTSM* without learning. Furthermore, agents will use fixed "hedge ratios" over time to manage the risks of their bond portfolios.[18]

[16]This is why the factor loadings B_m are reliably recovered from contemporaneous correlations among bond yields y_t^m and $\mathcal{P}_t$ (Duffee (2011)). It also explains why, holding $(K, N, Z, \mathcal{P})$ fixed, estimates of $\lambda^{\mathbb{Q}}$ is *DTSMs* without learning are virtually invariant across specifications of the $\mathbb{P}$ distribution of Z.

[17]In all subsequent figures, light green shaded areas correspond to NBER recessions, and the period between October, 1979 and October, 1982 represents the "Fed experiment" when the Federal Reserve focused on monetary aggregates instead of following interest rate rules.

[18]The estimates of the weights that define $\mathcal{P}$ are also stable over time.

This pattern in $\widehat{\lambda}_t^{\mathbb{Q}}$ is shared across all *DTSM*-based rules we examined.

Now yields also depend on the maturity-specific intercepts A_m, which in turn depend on agents' views about both $r_\infty^{\mathbb{Q}}$ (the long-run $\mathbb{Q}$-mean of r) and $\Sigma_{\mathcal{PP}}$. Though (as we will see) our $\mathcal{SC}$ agent substantially revises her estimates of $\Sigma_{\mathcal{PP}}$ over time, the impact of $\Sigma_{\mathcal{PP}}$ on A_m is through a convexity adjustment that is typically very small (see Appendix A). Therefore, by holding $\lambda^{\mathbb{Q}}$ fixed, our agent is effectively also fixing the A_m's. We conclude that our $\mathcal{SC}$ learning rules– which presumed that $\Theta^{\mathbb{Q}}$ is fixed and known– are not as naive as they might have seemed *ex ante*, as they produce pricing rules that are actually internally consistent.

Referring back to our comparative analysis of learning in Section 2.1, even with constant $\lambda^{\mathbb{Q}}$, $\mathcal{SC}$ learning under Case $\mathcal{B}\!\downarrow$**RLS** is not necessarily equivalent to full Bayesian learning for two reasons. First, the constraints on Bayesian updating on $\Theta^{\mathbb{P}}$ that would place it within Case $\mathcal{B}\!\downarrow$**ALS** may not hold. Second, with known $\lambda^{\mathbb{Q}}$, learning is mostly about $\Theta^{\mathbb{P}}$ and Σ_Z. The Bayesian treats Σ_Z explicitly as unknown, where as the $\mathcal{SC}$ agent computes her posterior distribution presuming that Σ_Z is known.

The relative accuracies of the rule-based forecasts, which depend primarily on $\widehat{\Theta}_t^{\mathbb{P}}$, can be assessed from the RMSE's displayed in Table 3. The RMSE's for forecasts of yields from the rules $\ell(BCFF)$ and $\ell(JSZ)$ (rule $\ell(JSZ_{CG})$ with $\gamma = 1$) are nearly the same for both one-quarter and one-year horizons, computed over the period January, 1985 through March, 2012. Rule $\ell(JSZ)$ out performs both of these rules for one-year forecasts for all maturities, and also for all of the yields beyond the six-month maturity over a one-quarter horizon. Below each RMSE are Diebold and Mariano (1995) (D-M) statistics for assessing whether two RMSE's are statistically the same, calculated as extended by Harvey, Leybourne, and Newbold (1997).[19] The first (in parentheses) tests against the $\ell(RM)$ rule, and the second (in brackets) tests against the rule implicit in $\ell(BCFF)$. The JSZ-based rules are not statistically different from the one used by the consensus professional at the one-year horizon, but they are statistically different over one-quarter horizons. By way of benchmarking, for the one-quarter forecast horizon $\ell(BCFF)$ performed about the same as the simple rule $\ell(RW)$; for one-year forecasts the RW model outperformed the consensus professional for longer maturity bonds.

A more nuanced view of relative forecast performance emerges from Figure 3 which compares the errors in forecasting $PC1$ and $PC2$ one-year ahead for the pair $\ell(BCFF)/\ell(JSZ)$.[20] While the errors from t he JSZ-based rule does not track those from $\ell(BCFF)$ perfectly, the tracking is remarkably close. The most challenging periods for matching $\ell(BCFF)$ by $\ell(JSZ)$ are the early 2000's (for slope), and the beginning of the recovery from the global financial crisis in 2011 (for level). Overall, the close match of rule $\ell(JSZ)$ to $\ell(BCFF)$ suggests that

[19] Consider two series of forecast errors e_{1t} and e_{2t}, $t = \{1, 2, ..., T\}$, define $d_t \equiv e_{1t}^2 - e_{2t}^2$, and let

$$\hat{\mu}_d = \frac{1}{T}\sum_{t=1}^{T} d_t \quad \text{and} \quad \hat{V}_d = \sum_{t=1}^{T}(d_t - \hat{\mu}_d)^2 + 2\sum_{j=1}^{h} k(j/h) \sum_{t=1}^{T-j}(d_t - \hat{\mu}_d)(d_{t+j} - \hat{\mu}_d),$$

where $k(.)$ is a Bartlett kernel that down-weights past lags to ensure that the variance of the difference in mean squared errors stays positive. The number of lags h is set to three for the one-quarter ahead forecasts and to twelve for the four-quarters ahead forecasts. Then D-M $= \sqrt{T}\hat{\mu}_d/\hat{V}_d^{1/2}$.

[20] The loadings for these PCs are normalized so that the forecast errors are expressed in basis points. Specifically, we scaled the loadings for $PC1$ to average out to unity, and the loadings for $PC2$ were scaled so that the difference between the loadings for the ten-year and six-month yields summed to unity.

Panel (a): RMSE's (in basis points) for Quarterly Horizon

Rule	6m	1Y	2Y	3Y	5Y	7Y	10Y
$\ell(RW)$	38.0	41.1	43.3	43.7	42.4	41.1	37.5
$\ell(BCFF)$	51.4	51.6	52.4	54.3	49.5	47.9	44.8
	()	()	()	()	()	()	()
	[4.10]	[3.28]	[4.48]	[5.03]	[4.86]	[3.40]	[3.54]
$\ell(JSZ)$	39.7	41.8	45.2	44.6	43.0	41.2	37.7
	(−4.03)	(−3.07)	(−3.92)	(−5.28)	(−4.39)	(−3.92)	(−3.33)
	[1.96]	[0.76]	[2.85]	[1.31]	[0.65]	[0.08]	[0.27]
$\ell(JSZ_{CG})$	38.5	41.6	45.2	45.0	43.4	42.1	38.8
	(−4.36)	(−3.17)	(−3.80)	(−4.45)	(−4.10)	(−3.66)	(−2.96)
	[0.50]	[0.48]	[3.05]	[1.55]	[1.20]	[1.21]	[2.01]
$\ell(JPS)$	36.2	41.2	44.2	43.9	41.4	40.7	39.3
	(−3.96)	(−2.74)	(−2.99)	(−3.86)	(−4.71)	(−3.94)	(−2.64)
	[−0.78]	[0.04]	[0.57]	[0.13]	[−1.20]	[−0.41]	[1.26]

Panel (b): RMSE's (in basis points) for Annual Horizon

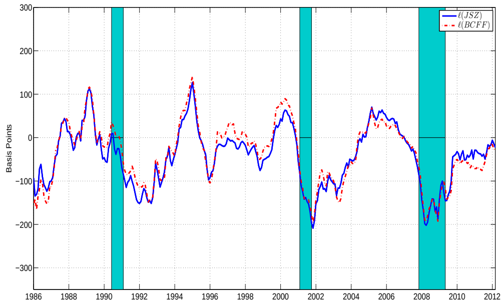
Rule	6m	1Y	2Y	3Y	5Y	7Y	10Y
$\ell(RW)$	136.2	135.3	126.3	118.0	107.3	102.2	96.0
$\ell(BCFF)$	148.2	144.6	140.1	136.2	119.6	113.9	106.0
	()	()	()	()	()	()	()
	[1.18]	[0.90]	[1.59]	[2.28]	[2.30]	[2.40]	[2.56]
$\ell(JSZ)$	141.7	140.6	134.7	125.9	111.7	102.3	92.9
	(−1.07)	(−0.51)	(−0.84)	(−1.61)	(−1.22)	(−1.66)	(−1.63)
	[0.75]	[0.77]	[1.26]	[1.28]	[0.81]	[0.02]	[−0.58]
$\ell(JSZ_{CG})$	137.3	136.6	130.5	122.5	110.7	104.1	97.4
	(−1.33)	(−0.92)	(−1.38)	(−1.93)	(−1.65)	(−1.85)	(−1.49)
	[0.19]	[0.26]	[0.92]	[1.01]	[1.14]	[0.72]	[0.50]
$\ell(JPS)$	130.4	130.7	123.3	114.4	101.8	96.5	92.8
	(−1.51)	(−1.31)	(−1.80)	(−2.52)	(−2.37)	(−2.23)	(−1.48)
	[−0.47]	[−0.42]	[−0.43]	[−0.72]	[−1.44]	[−1.12]	[−0.51]

Table 3: RMSE's for one-quarter (Panel (a)) and one-year (Panel (b)) ahead forecasts, January, 1985 to March, 2012. The D-M statistics for the differences between the *DTSM*- and BCFF-implied (*DTSM*- and *RW*-implied) forecasts are given in parentheses (brackets).

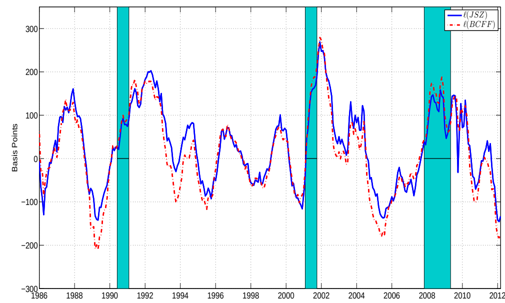
the consensus BCFF professional was effectively forecasting bond yields using a three-factor model with recursive least-squares updating of $\Theta^{\mathbb{P}}$.

The forecast errors are persistent, consistent with the long forecast horizon, and notably they tend to be large during NBER recessions when the consensus professional and $\ell(JSZ)$ forecasted much higher levels and steeper yield curves than what were experienced in the US. Also, after the Federal Reserve's shift away from their experimental monetary rule in the early 1980s, professionals predicted much higher rates than were realized for several years . This suggests that market participants were struggling with the credibility of the new monetary policy rule up until around 1987. We will explore in more depth some of the cyclical features of forecast errors in Section 6.1 where we take up the properties of risk premiums in the presence of learning with conditioning on macroeconomic information.

The RMSE's for rule $\ell(JSZ_{CG})$ which sets $\gamma = 0.99$ for down weighting the risk factors $\mathcal{P}$ (see Table 3) show virtually no improvement over those for $\ell(JSZ)$ over a quarterly



(a) $PC1 : \ell(BCFF)$ Versus $\ell(JSZ)$



(b) $PC2$: $\ell(BCFF)$ Versus $\ell(JSZ)$

Figure 3: Comparison of forecast errors (realization minus forecast) for $PC1$ and $PC2$ of rule $\ell(BCFF)$ against rule $\ell(JSZ)$ for the horizon of one year. The sample is January, 1985 through March, 2012.

forecast horizon. There is a larger gain for the one-year horizon (which was used to select γ), particularly for bonds with maturities under three years. These RMSE's are, however, misleading summaries of the degree to which the forecasts from $\ell(BCFF)$ and $\ell(JSZ_{CG})$ track each other. In fact, the tracking for both $PC1$ and $PC2$ is much inferior for $\ell(JSZ_{CG})$ than for $\ell(JSZ)$. This deterioration in tracking of BCFF forecasts comes with the benefit of much more accurate forecasts by $\ell(JSZ_{CG})$ than by either $\ell(BCFF)$ or $\ell(JSZ)$ during crisis, especially from mid-2009 onwards.

5 Bayesian Learning with Consistent Pricing

Our Bayesian learning rules bring greater sophistication to the consensus agent, a conceptual improvement over our naive learning rules, with some potential costs associated with added structure on agents' beliefs. The Bayesian agent specifies the joint distribution of her beliefs for a high-dimensional Θ and consistently updates these beliefs as new information becomes available. This requires specifying which parameters are known, which are unknown but constant, and which are drifting or state dependent, and for the latter parameters she must specify their laws of motion. With these beliefs in place, she then solves for prices that are consistent with the assumptions on parameter uncertainty.

Two learning schemes are explored in depth. In the first, Θ is presumed fixed over time and the consensus agent is learning its value. The second environment has the agent learning about the parameters governing the conditional $\mathbb{P}$-mean ($\Theta^{\mathbb{P}}$) as $\Theta^{\mathbb{P}}$ drifts according to the known law of motion (18)– a random walk reflecting permanent structural changes– presuming that the remaining parameters are fixed (but still unknown). Our interest in the second case is motivated in part by the debate in the macroeconomics literature about whether changes in monetary policies in the U.S. had material effects on the dynamics of VAR models of the macroeconomy (see, e.g., Cogley and Sargent (2005) and Sims and Zha (2006).)

Even for the simplest of these learning problems where all of the parameters are fixed there

	Panel (a): RMSE's (in basis points) for Three Months Horizon						
Model	6m	1Y	2Y	3Y	5Y	7Y	10Y
Fixed Param.	50.08	47.62	52.30	55.15	60.02	61.62	64.57
Drifting $\Theta^{\mathbb{P}}$	41.49	42.59	45.96	46.00	48.08	48.87	49.44
	Panel (b): RMSE's (in basis points) for Twelve Months Horizon						
Model	6m	1Y	2Y	3Y	5Y	7Y	10Y
Fixed Param.	195.82	187.55	181.59	176.48	169.00	162.97	158.22
Drifting $\Theta^{\mathbb{P}}$	162.52	158.92	139.59	149.23	128.47	123.39	119.04

Table 4: RMSE's for yield forecasts based Bayesian learning with the JSZ model over the period June, 1985 through March, 2012.

is not a closed-form solution for the joint posterior distribution of the parameters. Therefore, throughout we use the particle filter to approximate this distribution. We initialize parameter beliefs using maximum likelihood estimates over the sample between June, 1961 and June,1974. We then run the particle filter over the sample between June, 1974 and March, 2012. Owing to the highly modular structure of the particle filter, the recursions involved in solving models with drifting parameters are relatively straightforward extensions of the algorithm for the fixed-parameter case. Details are given in Appendix D.

Table 4 reports *RMSE*s for one-quarter and one-year ahead forecasts produced by the two learning algorithms set within the JSZ term structure setting (no conditioning on macroeconomic information). The Bayesian learner obtains more accurate forecasts, particularly over the one-year horizon, by adopting the prior that $\Theta^{\mathbb{P}}$ is drifting according to random walk. At the same time, notice that these RMSE's tend to be larger than the matching results for the relatively naive $\mathcal{SC}\ \ell(JSZ)$ rules (Table 3).

A likely explanation for the less accurate performance of the Bayesian learner is that we implemented these fully consistent learning rules under the assumption that the covariance matrix Q_t of the innovations in $\psi^{\mathbb{P}}$ (see (18)) is constant (not time dependent). In contrast, the updating under Case $\mathcal{B}\!\downarrow$**ALS** (which covers both *RLS* and *CGLS* learning), arises as a special case of Bayesian learning where Q_t is state-dependent and satisfies the constraints discussed in Section 2.1. The Bayesian agent formally accounts for the uncertain knowledge of $\Theta^{\mathbb{Q}}$ but, as we have seen, in our setting $\lambda^{\mathbb{Q}}$ is effectively known and approximately constant. This also diminishes the benefit of consistent pricing within the Bayesian rule, as the naive rules come close to internal consistency. The findings in Table 4 suggest that forecasts would be improved if we allowed the Bayesian agent to recognize state-dependence in Q_t.

6 Learning About Macroeconomic Risks and Monetary Policy

In this section we return to the properties of the *DTSM*-based learning rules, with particular emphasis on the evolution of the $\mathbb{P}$-parameters and the implied risk premiums as the consensus agent learns over time conditioning on both the shape of the Treasury curve and macroeconomic activity. Our learning window begins in January, 1985 shortly after the return to an interest-rate policy rule by the Federal Reserve Board (FRB). In the early 1990's the first Iraq war led to a mild recession, and over this period the FRB progressively cut the policy rate from 8% at the beginning of the Gulf war to 6% in the spring of 1991. The federal funds rate was again cut to 3% in 1992 and remained at this level until February, 1994. Since inflation was close to 3% at the time, the short-term real rate was essentially zero.

There were two particularly relevant developments in the early 1990's for our analysis of learning. In 1994 the US economy was in a strong expansion and the FRB, in order to preemptively counteract increases in nominal prices, tightened monetary policy resulting (by February, 1995) in an increase in bond yields of 2.5% and a substantial steepening of the yield curve. Additionally, starting in February, 1994 the FRB started announcing their federal funds rate target immediately after each FOMC meeting, thereby making policy communication more transparent. Similarly, in early 1999 the FRB started immediately communicating major changes in its views on the direction of monetary policy. Wright (2011) and Bauer, Rudebusch, and Wu (2013) attribute the declines in bond-market risk premiums in the 1990's to increased transparency and declining uncertainty about inflation.

For exploring the links between macroeconomic activity and learning rules, we focus on the more flexible and informationally rich learning rule $\ell(JPS)$. The RMSE's for the $\ell(JPS)$ rule over our entire learning period from 1985 to 2012 (Table 3) reveal that professionals could have improved their forecasts by incorporating information about inflation and output growth in the manner it enters the JPS model. The gain in forecast accuracy over $\ell(BCFF)$ is statistically significant for bonds of all maturities except the ten-year over the one-quarter horizon and borderline significant for the intermediate maturity bonds when forecasting one-year ahead.[21]

Table 5 compares one-year-ahead RMSE's for the three sub-periods January, 1985 through December, 1999, roughly the first half of our learning period and ending about the time of several major changes in FRB disclosure policies; January, 2000 through December, 2007, the period of asset price inflation leading up to the global financial crisis; and January, 2008 through March, 2012 covering the crisis up to the end of our sample. Notably, the RMSE's for the first half of our sample tend to be much larger than those for the crisis period, and they are more comparable to the full-sample results in Table 3 Panel (b). Apparently market participants and the *DTSM*-based rules alike, when faced with selloffs in financial markets and the aggressive non-standard policies implemented by the central banks globally during the crisis, could relatively reliably predict the future paths of bond yields.

The most challenging period for predicting bond yields was during the early 2000's, starting with the dotcom bust and then leading up to the recent global crisis. The gain in forecasting accuracy from conditioning on M is particularly large during this period, as rules $\ell(BCFF)$ and $\ell(JPS)$ give very different RMSE's. The difficulty in forecasting was particularly acute for short maturity bonds (under five years to maturity). That conditioning on inflation and

[21]Throughout our analysis of the rule $\ell(JPS)$ we do not enforce the constraint that the largest (most persistent) eigenvalues of the feedback matrices $K_Z^{\mathbb{P}}$ and $K_Z^{\mathbb{Q}}$ being equal, as was done in JPS's preferred model. This constraint is interesting in our learning setting, because it effectively has the agent using the precise knowledge of $\lambda^{\mathbb{Q}}$ from the cross-section to mitigate small-sample bias in estimating the parameters $K_Z^{\mathbb{P}}$ from the $\mathbb{P}$-Markov process for f. For similar reasons Jardet, Monfort, and Pegoraro (2012) enforced near-cointegration in their *DTSMs*. Nevertheless, we find that enforcing this constraint add little to improving out-of-sample forecasts from our learning rules.

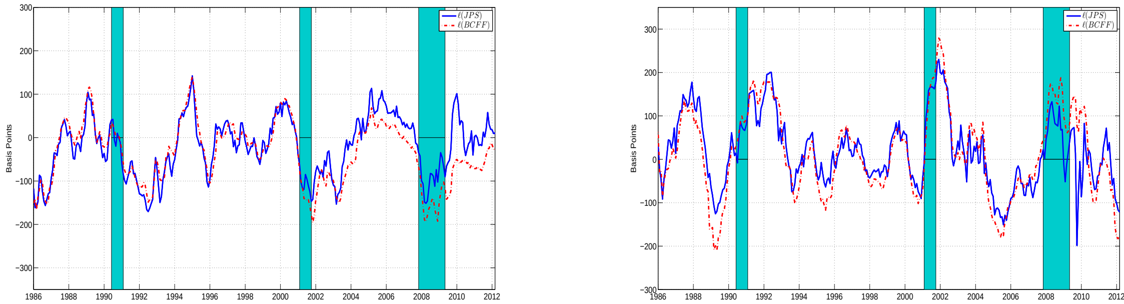
Rule	RMSE's by Bond Maturity 6m	1Y	2Y	3Y	5Y	7Y	10Y
	January, 1985 – December, 1999						
$\ell(RW)$	124	127	127	123	117	117	115
$\ell(BCFF)$	137	138	133	129	121	120	120
$\ell((JSZ)$	124	127	124	120	111	108	106
$\ell((JSZ_{CG})$	130	134	131	127	119	117	114
$\ell((JPS)$	132	135	130	124	114	111	107
	January, 2000 – December, 2007						
$\ell(RW)$	173	165	143	125	98	79	60
$\ell(BCFF)$	178	165	156	144	116	98	79
$\ell(JSZ)$	181	176	163	145	118	97	75
$\ell(JSZ_{CG})$	166	159	145	128	104	86	69
$\ell(JPS)$	141	138	125	109	86	71	64
	January, 2008 – March, 2012						
$\ell(RW)$	75	75	67	67	76	78	69
$\ell(BCFF)$	116	118	129	148	122	119	94
$\ell(JSZ)$	100	97	102	103	98	85	67
$\ell(JSZ_{CG})$	78	76	76	79	82	79	71
$\ell(JPS)$	92	87	79	75	77	76	78

Table 5: RMSE's in basis points for one-year-ahead forecasts over the indicated periods.

output growth is what substantially improved the performance of $\ell(JPS)$ is apparent from higher RMSE's for $\ell(JSZ)$ out to the seven-year segment of the Treasury curve. Even rule $\ell(JPS_{CG})$, which was optimized by choice of $\gamma = 0.99$ for the entire sample period, shows much weaker forecasting accuracy relative to $\ell(JPS)$ during this period. Additionally, $\ell(JPS)$ also substantially outperformed $\ell(RW)$ for the shorter maturity bonds.

It is our impression that during this period many central banks' used *DTSMs* that conditioned only on the shape of the yield curve when extracting estimates of market risk premiums and, as such, they were using suboptimal forecasting models. Figure 4(a) shows the systematic over-forecasting of the level of yields by both professionals and the yield-only rule $\ell(JSZ)$ (predicted yields much higher than those subsequently realized) for most of this period. Starting in the depths of the 2001 recession, $\ell(JPS)$ outperformed $\ell(BCFF)$ through 2004. Only during 2005 - 2006, when these learning rules systematically under-forecasted yields, did $\ell(BCFF)$ outperform $\ell(JPS)$.

Another period during which the professionals had a relatively difficult time forecasting bond yields was post 2008, during the crisis. However, in this case it was not so much conditioning on macroeconomic information that helped the *DTSM*-based rules, but rather the smooth updating implicit in $\mathcal{SC}$ learning. At the beginning of the crisis $\ell(JPS)$ systematically out performs $\ell(BCFF)$, because the professionals were consistently, and incorrectly, predicting that Treasury bond yields would rise. The gradual *RLS* (*CGLS*) updating within $\ell(JPS)$



(a) $PC1$: $\ell(BCFF)$ Versus $\ell(JPS)$ (b) $PC2$: $\ell(BCFF)$ Versus $\ell(JPS)$

Figure 4: Comparison of forecast errors (realization minus forecast) for $PC1$ and $PC2$ of rule $\ell(BCFF)$ against rule $\ell(JPS)$ for the horizon of one year. The sample is January, 1985 through March, 2012.

($\ell(JSZ_{CG})$) was more consistent with what actually transpired. This is also apparent in Figure 4(a) for rule $\ell(JPS)$. This smooth updating also explains the spike for $\ell(JPS)$ in forecast errors for $PC1$ around the turn of the year 2010. Ten-year Treasuries reached a trough of 2.4% in December, 2008. Rule $\ell(JPS)$ extrapolated the low bond yields ahead to the end of 2009. What happened in fact was that yields rose more rapidly, finally validating the earlier (incorrect) views of $\ell(BCFF)$ and leading to substantial under prediction by $\ell(JPS)$.

The three- to five-year segment of the Treasury curve was a particularly challenging segment for professional forecasters during the crisis, strikingly so relative to either of the earlier subperiods. This segment of the curve is intriguing because Fleming and Remolona (1999) show that macroeconomic announcements had a relatively large impact on the shape of the intermediate segment of the yield curve. Additionally, Piazzesi (2005)'s fitted monetary policy rule suggests that the FRB reacted to information in the two-year segment of the Treasury curve. These connections between changing shapes of the Treasury curve and macroeconomic activity may have been more accurately captured in $\ell(JSZ)$ and $\ell(JPS)$ than in $\ell(BCFF)$. A complementary explanation for the poor forecasting performance by professionals relates to the roles of intermediate maturity bonds in hedging mortgage-related securities (Duarte (2008)). The collapse of the MBS market may also have been a factor and, if so, it seems to have affected professionals much more so than *DTSM*-based learners.

The best performing rule during the crisis (for this one-year horizon) is $\ell(RW)$, though rule $\ell(JSZ_{CG})$ produces inconsequentially larger RMSE's. That constant-gain learning absent conditioning on macroeconomic information (slightly) outperforms least-squares learning with conditioning on M serves to highlight the importance of an adaptable learning rule that was responsive to major changes in policy during this period.

Of equal interest is how the various learning rules forecasted $PC2$. The slope of the yield curve is often linked directly to the stance of monetary policy (e.g., Rudebusch and Wu (2008)) and $PC2$ has been found to have strong predictive power for risk premiums in Treasury markets (e.g., Cochrane and Piazzesi (2005)). Recall from Figure 3(b) that forecasts of $PC2$

from $\ell(JSZ)$ tracked those of $\ell(BCFF)$ quite closely over our entire sample. In contrast, from Figure 4(b) it is seen that forecasts of $PC2$ from $\ell(JPS)$ deviate systematically from those of $\ell(BCFF)$ throughout the sample.

Moreover, when there deviations are large, it is rule $\ell(JPS)$ that is much more accurate. Particularly during periods when forecasts of PC2 were well above the subsequently realized slopes of the yield curve (the troughs in Figure 4(b)), $\ell(JPS)$ outperformed $\ell(BCFF)$. It turns out that these discrepancies happen around local turning points for $PC2$, when the yield curve transitioned from steepening to flattening and vice versa.

6.1 Risk Premiums in the Presence of Learning

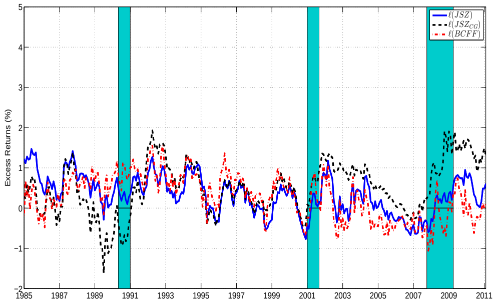
Within the literature on reduced-form *DTSMs*, risk premiums have typically be computed under the presumption that market participants have full knowledge of the laws of motion of the risk factors (Duffee (2001), Dai and Singleton (2002)). In the presence of learning, there may be variation in the $\mathcal{SC}$ consensus agent's expected excess returns induced by the learning process *per se* as contrasted with variation induced by movement in the underlying quantities or market prices of the risks $\mathcal{P}_t$. In this section we quantify this learning component of risk premiums, and relate its properties to the evolution of the macroeconomy.[22]

Figure 5 displays the *expected* excess returns over one-year holding periods for positions in two- and ten-year bonds for three learning rules: $\ell(BCFF)$, $\ell(JSZ)$, and $\ell(JSZ_{CG})$. We focus on yields on individual bonds so as to avoid the rebalancing and approximations involved in computing multi-period expected excess returns for the PC-mimicking portfolios. In the case of the two-year bond the three rule-implied risk premiums track each other quite closely for most of the sample. The relatively rapid adjustment inherent in the $CGLS$-based rule $\ell(JSZ_{CG})$ induces large differences with the other two rules in the late 1980's and after 2001. The sensitivity to rule-based risk premiums to the choice of γ is also striking. While RMSE's for forecasts differed somewhat for $\ell(JSZ)$ and $\ell(JSZ_{CG})$, those results did not foreshadow the very distinctive patterns in Figure 5(a).

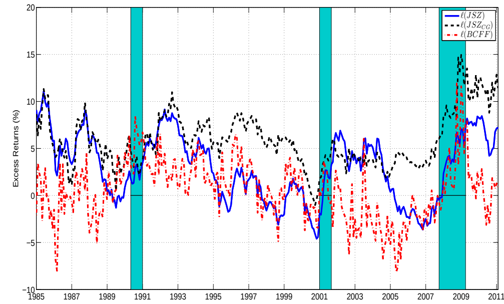
For most of our sample period the compensations for bearing two-year Treasury bond risk were positive, with only occasional and short-lived dips below zero. The primary exception is the period from the beginning of 2005 through 2007 when risk premiums were systematically declining and turned negative. This is the period that Chairman Greenspan referred to as the conundrum, because long-term bond yields remained largely unchanged while the FRB was raising short-term policy rates.

Whereas risk premiums on positions in two-year bonds are broadly similar across the learning rules, they are notably different for positions in ten-year bonds (Figure 5(b)). Rules $\ell(JSZ)$ and $\ell(BCFF)$ roughly track each other in the late 1980's, the second half of the 1990's, and during 2007-08, but otherwise they are very different. The risk premiums implied by $\ell(JSZ_{CG})$ are distinctive for virtually the entire post-1994 period. Varying the down weighting of past data through the choice of γ has huge effects on the risk premiums on

[22]Focusing on learning is complementary to, but distinct from variation induced by ambiguity aversion (e.g., Barillas, Hansen, and Sargent (2009), Ulrich (2013)). An interesting question for future research is how learning and ambiguity about the correct underlying model of factor risks interact within a term structure setting. Hansen (2007) explores these interactions at a broad macroeconomic level.



(a) Two-Year Bond



(b) Ten-Year Bond

Figure 5: Average expected excess returns over holding periods of ten, eleven and twelve months for the two- and ten-year bonds based on learning rules $\ell(JSZ)$, $\ell(JSZ_{CG})$, and $\ell(BCFF)$, January, 1985 to March, 2011.

long-term bonds implied by learning rules.

Pursuing the differences between $\ell(JSZ)$ and $\ell(BCFF)$, notice that shortly after every NBER recession in our sample the gaps between the risk premiums implied by these rules become relatively large. As can be seen from Figure 6, a key factor underlying these differences is the slope of the yield curve. The strong positive correlation between compensation for bearing the risk of holding ten-year bonds and the steepness of the yield curve is evident. Moreover, it is precisely when the Treasury curve is relatively very steep that the consensus professional forecaster believed that risk compensation was much lower than what was implied by our *DTSM*-based learning rule.

To translate these findings into implications for views on the shape of the Treasury curve, consider the period from 2001 through 2005. The yield y^{10y} reached a local peak of about 6.7% in early January, 2000, and it fell to a local low of about 3.7% at the beginning of October, 2002. It then bounced around in a moderately narrow trading range and ended 2005 at about 4.4%. The $y^{10y} - y^{2y}$ spread was inverted at the beginning of 2000 and it steepened substantially through the middle of 2004. The difference between the expected excess returns from $\ell(JSZ)$ and $\ell(BCFF)$ arises entirely from different forecasts of the ten-year yield one year ahead. As discussed above, the professional forecasters tended to expect a more rapid rise in long-term Treasury yields relative to the forecasts implied by $\ell(JSZ)$ in the periods following recessions. It is these differences that underlie the relatively higher risk compensation demanded by the agent following the learning rule $\ell(JSZ)$.

Among the agents following $\ell(JSZ)$ or $\ell(BCFF)$, which one had more accurate assessments of realized excess returns? Over the entire sample period, the RMSE's in forecasting the realized excess returns for bearing $(2y, 10y)$ bond risks were $(1.55\%, 9.68\%)$ for $\ell(BCFF)$ and $(1.47\%, 7.97\%)$ for $\ell(JSZ)$. For this specific episode over January, 2001 through January, 2006, the corresponding RMSE's were $(1.34\%, 7.62\%)$ for $\ell(BCFF)$ and $(1.40\%, 4.34\%)$ for $\ell(JSZ)$. The two rules had comparable forecast accuracy for risk compensation on the two-year bond. On the other hand, particularly for the post-recession periods, the out-performance of $\ell(JSZ)$

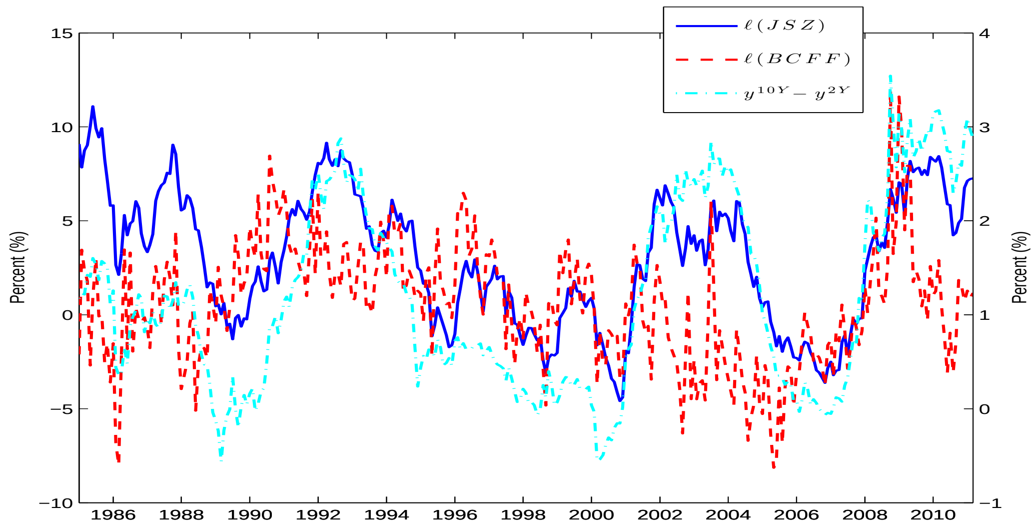


Figure 6: Average expected excess returns over holding periods of ten, eleven and twelve months for the ten-year bond based on $\ell(JSZ)$ and $\ell(BCFF)$ (left axis) and the slope of the Treasury curve measured as $y^{10y} - y^{2y}$ (right axis), January, 1985 to March, 2011.

in forecasting excess returns on the ten-year bond was enormous. We stress that rule $\ell(JSZ)$ is a fully *ex ante* learning rule that could have been well approximated throughout our sample by a factor-VAR forecasting model, and rule $\ell(JSZ)$ was available to market participants during the recent crisis when it again substantially out performed $\ell(BCFF)$.

Providing an economic interpretation of these differences in challenging absent a model that reveals whether the BCFF professionals were marginal traders in Treasury bonds. If these professionals were marginal, then the prices of Treasuries would likely have been substantially different had they been learning by $\ell(JSZ)$. On the other hand, if they were not marginal and were simply following inefficient learning rules, then a key message from our findings is that disciplining *DTSMs* using survey data on long-term yields will likely lead to distorted measures of required risk compensations.[23]

With this caveat in mind, it is of interest to examine the economic factors underlying the discrepancies between the risk premiums implied by $\ell(BCFF)$ and $\ell(JSZ)$. Table 6 displays the R^2's from the projections of rule-implied *expected* excess returns (ER) onto various sets of conditioning variables. Note, first of all, that roughly 15% of the variation in $ER_2[\ell(JSZ)]$ is unspanned by $\mathcal{P}$; the conventional *DTSM* without learning implies complete (100%) spanning. It follows that roughly 15% of the variation in $ER_2[\ell(JSZ)]$ is a pure learning effect. The learning effect is much weaker for ten-year bonds, as the corresponding R^2 is 94%. Second, adding M to the conditioning information has a negligible effect on the R^2, indicating that the learning effect is not linked directly to developments with inflation

[23]Gains in forecast performance may come from using information embedded in survey forecasts of short-term rates and, indeed, Altavilla, Giacomini, and Ragusa (2014) present evidence consistent with this view.

Rule	January, 1987 Through March, 2011					
	Excess Return 2 Year Bond					
Rule	$\mathcal{P}$	$\mathcal{S}_{000}$	$\mathcal{S}_{100}$	$\mathcal{S}_{101}$	$\mathcal{S}_{011}$	$\mathcal{S}_{001}$
$\ell(JSZ)$	84.0	89.7	91.1	91.1	89.8	89.8
$\ell(BCFF)$	71.7	72.0	73.2	73.4	75.4	72.6
$\ell(JSZ) - \ell(BCFF)$	28.9	36.7	46.9	47.3	46.5	39.4
	Excess Return 10 Year Bond					
Rule	$\mathcal{P}$	$\mathcal{S}_{000}$	$\mathcal{S}_{100}$	$\mathcal{S}_{101}$	$\mathcal{S}_{011}$	$\mathcal{S}_{001}$
$\ell(JSZ)$	94.1	95.5	97.0	97.0	95.7	95.5
$\ell(BCFF)$	22.3	37.0	41.0	51.6	54.9	50.8
$\ell(JSZ) - \ell(BCFF)$	44.7	51.4	59.6	66.3	67.4	62.5

Table 6: R^2's from projections of learning rule-implied *expected* excess returns for the one-year holding period. The conditioning sets $\mathcal{S}_{ijk}$ indicate the combination of the yield PCs ($\mathcal{P}$), macroeconomic factors (M), the liquidity factor (LIQ), and $i : MAD(INF)$, $j : MAD(PC1)$, $k : MAD(PC2)$.

or output growth. We also conditioned on the liquidity variable (LIQ) constructed for Treasury markets by Hu, Pan, and Wang (2013), and it also added no explanatory power for $ER_2[\ell(JSZ)]$. Finally, returning to the theme of heterogeneity discussed in Section 2.2, we examined whether variation in $ER_2[\ell(JSZ)]$ was correlated with the forecast dispersion measures $(MAD(INF), MAD(PC1), MAD(PC2))$ indicated by the indicators (i, j, k) on $\mathcal{S}_{ijk}$. From the first row of Table 6 the answer appears to be no. The results for $ER_2[\ell(BCFF)]$ are similar, though the corresponding R^2's are consistently a bit lower. Were one to view $ER_2[\ell(BCFF)]$ as being generating by a *DTSM*, then it would follow that the learning effect is more pronounced for the professional forecasters.

Of particular interest is whether any of these conditioning variables are significantly correlated with the difference $ER_2[\ell(BCFF)] - ER_2[\ell(JSZ)]$ which, as we have noted, becomes especially large shortly after every economic downturn since 1985. This gap is highly correlated with the PCs, LIQ and $MAD(INF)$, with comparable explanatory power being achieved by replacing $MAD(INF)$ with $(MAD(PC1), MAD(PC2)$.[24] For the case of the ten-year bond, the dispersion variables again have substantial incremental explanatory power for the difference between the returns. Conditioning on either $(MAD(INF), MAD(PC2))$ or $(MAD(PC1), MAD(PC2))$ leads to projections that account for two-thirds of the variation in $ER_{10}[\ell(BCFF)] - ER_{10}[\ell(JSZ)]$.

Viewing these patterns together, it appears as though the consensus forecasts of bond yields reflect the heterogeneous views among the professional forecasters about the future course of inflation (or $PC1$) and the slope of the yield curve. To the extent we can connect uncertainty about the slope of the Treasury curve to uncertainty about the future course of monetary policy, it appears as though lack of agreement about inflation and monetary policy

[24] The forecast dispersion series are highly (though far from perfectly) correlated over our sample period. They tended to start high in the late 1980's, to mildly cycle during the 1990's and 2000's, and then $MAD(INF)$ spikes up during the crisis (much more so than $MAD(PC1)$ or $MAD(PC2)$).

sent the median professional off course from efficient forecasting relative to the *DTSM*-based rule $\ell(JSZ)$. Additionally, it seems notable that uncertainty about inflation (as measured by $MAD(INF)$) has so much explanatory power for the under-performance of the professional forecasters, even at the intermediate two-year point on the Treasury curve. The accounting in Table 6 is conservative in the sense that $\mathcal{P}$ is also highly correlated with the MAD's (see Table 1) and the conditioning sets $\mathcal{S}_{ijk}$ include $\mathcal{P}$.

6.2 The Evolution of Beliefs About Parameters and Monetary Policies

Not surprisingly, within our sample period, the largest and most frequent revisions in $(K_Z^{\mathbb{P}}, \Sigma_{\mathcal{P}\mathcal{P}})$ occurred during the Fed experiment between the falls of 1979 and 1982. During this period of a monetary-base focused policy, there was a sharp rise in interest rates and flattening of the Treasury curve. Figure 7 displays the consensus agent's views– as implied by the learning rule $\ell(JSZ_R)$– about the unconditional covariance matrix of the risk factors $\mathcal{P}$ and Figure 8 shows the revisions in $K_{\mathcal{P}}^{\mathbb{P}}$.

There is a notable and sharp increase in the perceived innovation variances of the yield PCs at the beginning of the FRB experiment. Interestingly, whereas the perceived risks of $PC1$ and $PC3$ started to gradually decline after the announced return to an interest-rate targeting rule, the perceived risk of the slope factor $PC2$ remained high until 1985. Concurrently, during the early 1980's, there was considerable instability in the agent's views about the mean reversion parameters $K_{\mathcal{P}}^{\mathbb{P}}$. We find it interesting that the consensus agent attempts to match the perceived increases in the volatilities of bond yields at the beginning of the FRB experiment by increasing both her perceived unconditional variance of $\mathcal{P}_t$ and the degrees of persistence of these factors. This is particularly evident for the parameters governing the dynamics of $PC1$, the first entry of $\mathcal{P}$. In fact, the diagonal element of $K_{\mathcal{P}}^{\mathbb{P}}$ for $PC2$ actually declines during most of the period of the Fed experiment.

From Figures 7 and 8 it appears these developments had imperceptible effects on the one-month conditional covariance matrix of $\mathcal{P}$. A more nuanced view emerges from the paths of the conditional correlations among the $\mathcal{P}$ based on the rule $\ell(JKPS)$ (see Figure 9).[25] The correlation between the level and slope factors spiked to 0.60 during the Fed experiment and it remained high well into 1985, at which time it began a gradual return to near 1970's levels.

[25] Recall that the loadings defining these PCs that emerge from our learning rules were virtually constant over our sample period, so any changes in correlations were not induced by agents changing the measurement of the PCs themselves.

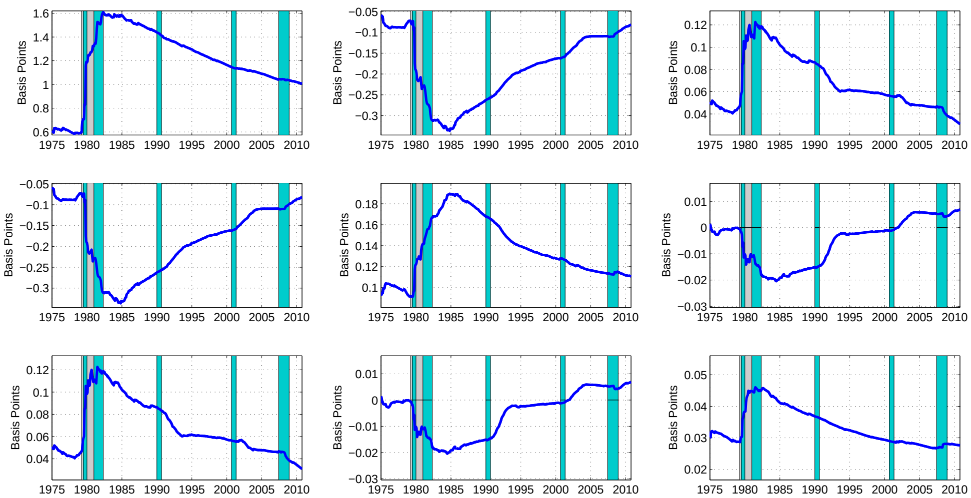


Figure 7: Estimates from $\ell(JPS)$ of $\Sigma^{\mathbb{P}}_{\mathcal{P}\mathcal{P}}$, the innovation covariance matrix for $\mathcal{P}_t$, over the period June, 1975 to March, 2011.

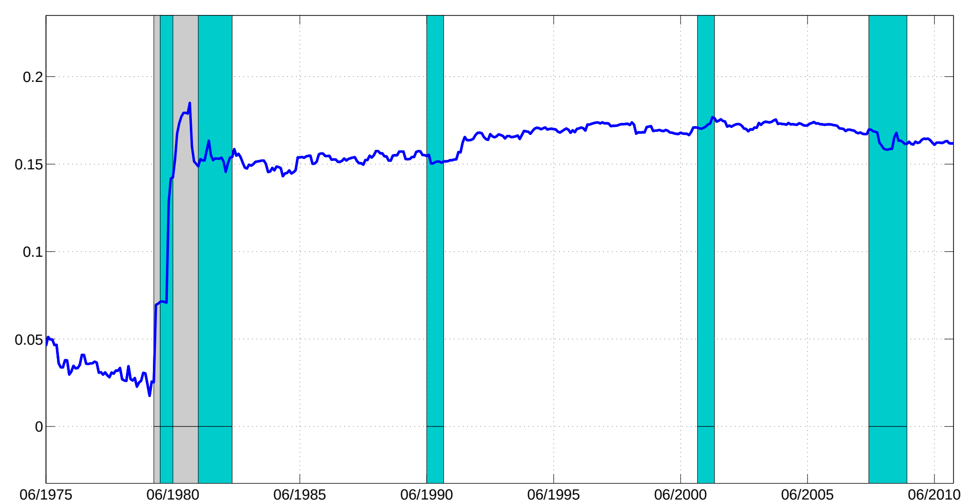


Figure 10: The *conditional* correlation between $\mathcal{P}_1$ and NAI, over the period June, 1975 to March, 2011, based on the learning rule $\ell(JSZ)$.

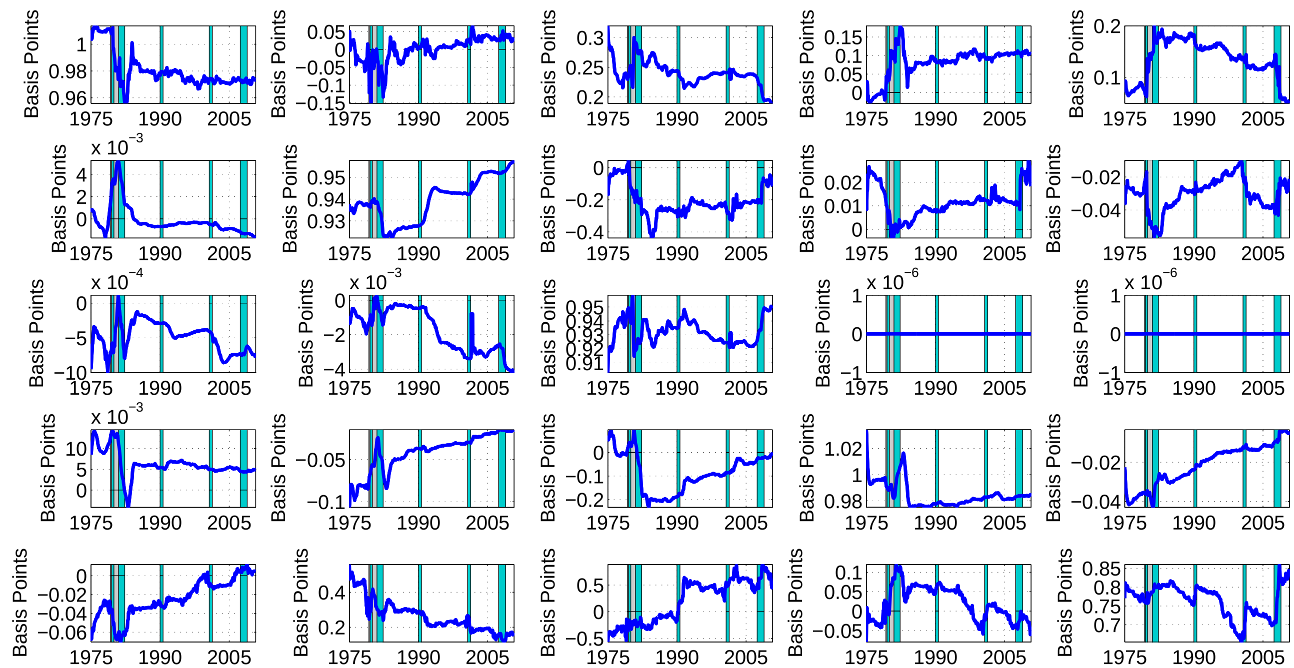


Figure 8: Estimates from $\ell(JPS)$ of the feedback matrix $K_Z^{\mathbb{P}}$ over the period June, 1975 to March, 2011.

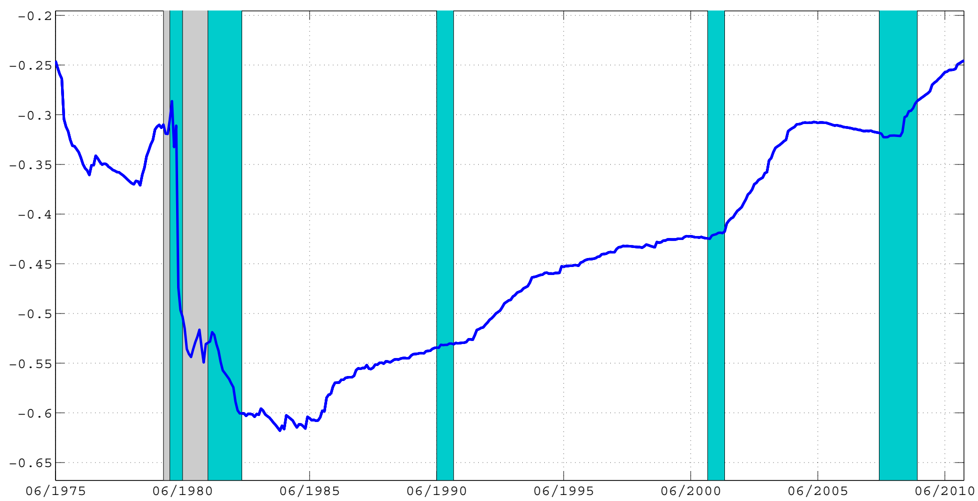


Figure 9: The *conditional* correlation between $\mathcal{P}_1$ and $\mathcal{P}_2$, over the period June, 1975 to March, 2011, based on the learning rule $\ell(JSZ_R)$.

7 Dispersion of Beliefs and Bond Market Risk Premiums

Of particular interest to us is whether the $MAD(PCi)$ have explanatory power for expected excess returns on bond portfolio positions beyond the information in the core risk factors $\mathcal{P}_t$. For, were this the case, then our learning framework based on a consensus agent and the conditioning information $Z = (\mathcal{P}, M)$ would be omitting information that models with heterogeneity suggest is important. Table 7 displays the projections of the realized excess returns for one-year holding periods on bonds of various maturities onto Z and the MAD dispersion measures for $(\mathcal{P}, INF)$.

8 Conclusions and Further Work

	2Y	3Y	5Y	7Y	10Y
Const.	−0.0056	−0.0056	−0.0069	−0.0075	−0.0115
	(0.0055)	(0.0114)	(0.0207)	(0.0275)	(0.0388)
	[−1.0171]	[−0.4940]	[−0.3338]	[−0.2731]	[−0.2966]
$PC1$	0.5962	1.1949	2.3288	3.8639	5.9045
	(0.2156)	(0.4322)	(0.7888)	(1.0719)	(1.5576)
	[2.7650]	[2.7647]	[2.9523]	[3.6046]	[3.7908]
$PC2$	0.0175	0.0501	0.1463	0.2761	0.4675
	(0.0157)	(0.0288)	(0.0535)	(0.0719)	(0.1037)
	[1.1203]	[1.7379]	[2.7365]	[3.8394]	[4.5079]
$PC3$	0.0375	0.0777	0.1854	0.3254	0.5204
	(0.0269)	(0.0551)	(0.1035)	(0.1378)	(0.1975)
	[1.3936]	[1.4100]	[1.7923]	[2.3615]	[2.6347]
$Infl$	−0.5253	−1.5187	−4.0224	−7.4737	−12.6084
	(0.4900)	(0.9728)	(1.7769)	(2.3944)	(3.3718)
	[−1.0720]	[−1.5612]	[−2.2638]	[−3.1213]	[−3.7394]
NAI	−0.0084	−0.0158	−0.0213	−0.0274	−0.0413
	(0.0032)	(0.0059)	(0.0104)	(0.0130)	(0.0160)
	[−2.6481]	[−2.6651]	[−2.0413]	[−2.1006]	[−2.5860]
$MAD(PC1)$	−0.0071	−0.0129	−0.0223	−0.0286	−0.0325
	(0.0024)	(0.0047)	(0.0090)	(0.0124)	(0.0179)
	[−2.9452]	[−2.7547]	[−2.4751]	[−2.3018]	[−1.8139]
$MAD(PC2)$	−0.0022	−0.0029	−0.0013	0.0035	0.0118
	(0.0022)	(0.0042)	(0.0082)	(0.0114)	(0.0168)
	[−1.0212]	[−0.6804]	[−0.1586]	[0.3083]	[0.7049]
$MAD(PC3)$	0.0079	0.0138	0.0207	0.0222	0.0203
	(0.0024)	(0.0044)	(0.0077)	(0.0105)	(0.0149)
	[3.3151]	[3.1523]	[2.7005]	[2.1157]	[1.3671]
$MAD(Infl)$	−0.0010	−0.0014	0.0003	−0.0007	−0.0009
	(0.0019)	(0.0039)	(0.0075)	(0.0100)	(0.0135)
	[−0.5542]	[−0.3669]	[0.0464]	[−0.0708]	[−0.0682]
Adj. R^2	0.3777	0.3187	0.2817	0.3185	0.3519
WALD	15.0426	13.0518	9.6068	7.1557	3.8941

Table 7: Predictive regressions for one year excess returns of bonds with maturities of 2, 3, 5, 7 and 10 years. Robust standard errors are in parentheses. WALD denotes the Wald statistic for the null hypothesis that the coefficients on the three MAD variables are equal to zero. The 95th percentile from the distribution of the Wald statistic under the null is equal to 9.4877. The sample period is January,1985 through March, 2012.

A Bond Pricing in *GTSMs*

The price of a zero coupon bond is given by:

$$D_{t,m} = e^{\mathcal{A}_m + \mathcal{B}_m \mathcal{P}_t}$$

Where A_m and B_m solve the first order difference equation:

$$\mathcal{A}_{m+1} - \mathcal{A}_m = \left(K_0^Q\right)' \mathcal{B}_m + \frac{1}{2}\mathcal{B}_m' \Sigma_{\mathcal{P}} \mathcal{B}_m - \rho_0$$

$$\mathcal{B}_{m+1} - \mathcal{B}_m = \left(K_1^Q - I\right)' \mathcal{B}_m - \rho_1$$

With initial conditions $\mathcal{A}_0 = 0$ and $\mathcal{B}_0 = 0$. The corresponding loadings for yields will be $A_m = -\mathcal{A}_m/m$ and $B_m = -\mathcal{B}_m/m$.

B The Canonical Model

We devote this appendix to the derivation of the canonical model in JSZ 2011. We start from the the equations for the dynamics of the factors and the short rate:

$$\begin{aligned} X_{t+1} &= J(\lambda^{\mathbb{Q}})X_t + \Sigma_X^{1/2} e_{X,t+1} \\ r_t &= r_\infty^{\mathbb{Q}} + 1 \cdot X_t \end{aligned}$$

Note that also in this framework we can solve for bond prices using the recursion:

$$D_{m+1,t} = E^{\mathbb{Q}}\left[E^{\mathbb{Q}} D_{m,t+1} \exp(-r_t)\right]$$

We will obtain that:

$$y_{m,t} = A_{X,m}\left(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_X\right) + B_{X,m}\left(\lambda^{\mathbb{Q}}\right) X_t$$

With Riccati equations:

$$\begin{aligned} A_{X,m+1} - A_{X,m} &= \frac{1}{2} B_{X,m}^{'} \Sigma_X B_{X,m} - r_\infty^{\mathbb{Q}} \\ B_{X,m+1} - B_{X,m} &= J\left(\lambda^{\mathbb{Q}}\right)^{'} B_{X,m} \end{aligned}$$

We now assume that we can exactly rotate pricing factors into portfolios. Then:

$$\begin{aligned} \mathcal{P}_t &= W A_{y,X}(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_X) + W B_{y,X}(\lambda^{\mathbb{Q}}) X_t \\ \mathcal{P}_t &= C(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_X) + D(\lambda^{\mathbb{Q}}) X_t \\ X_t &= D(\lambda^{\mathbb{Q}})^{-1}\left(\mathcal{P}_t - C(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_X)\right) \end{aligned}$$

Where W is a matrix of eigenvectors obtained from the eigendecomposition of the variance-covariance matrix of yields. We can now calculate the conditional Q expectations of the principal components:

$$\begin{aligned} E_t^{\mathbb{Q}}\left[\mathcal{P}_{t+1}\right] &= C(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_X) + D(\lambda^{\mathbb{Q}}) E_t^{\mathbb{Q}}\left[X_{t+1}\right] \\ &= C(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_X) + D(\lambda^{\mathbb{Q}}) J\left(\lambda^{\mathbb{Q}}\right)^{'} D(\lambda^{\mathbb{Q}})^{-1}\left(\mathcal{P}_t - C(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_X)\right) \\ &= C(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_X) - D(\lambda^{\mathbb{Q}}) J\left(\lambda^{\mathbb{Q}}\right)^{'} D(\lambda^{\mathbb{Q}})^{-1} C(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_X) + D(\lambda^{\mathbb{Q}}) J\left(\lambda^{\mathbb{Q}}\right)^{'} D(\lambda^{\mathbb{Q}})^{-1} \mathcal{P}_t \end{aligned}$$

And their conditional variance as:

$$\begin{aligned} Var_t^{\mathbb{Q}}\left[\mathcal{P}_{t+1}\right] = Var_t^{\mathbb{Q}}\left[\mathcal{P}_{t+1}\right] &= D(\lambda^{\mathbb{Q}}) Var_t^{\mathbb{Q}}\left[X_{t+1}\right] D(\lambda^{\mathbb{Q}}) \\ &= D(\lambda^{\mathbb{Q}}) \Sigma_X D(\lambda^{\mathbb{Q}}) \end{aligned}$$

We can then finally rewrite the term structure model using the principal components as the state variables:

$$\begin{aligned} \mathcal{P}_{t+1} &= K_0^{\mathbb{Q}} + K_1^{\mathbb{Q}} \mathcal{P}_t + \Sigma_{\mathcal{P}}^{1/2} e_{\mathcal{P},t+1} \\ r_t &= \rho_0 + \rho_1 \mathcal{P}_t \end{aligned}$$

Where:

$$
\begin{aligned}
K_0^Q &= -D(\lambda^{\mathbb{Q}})J(\lambda^{\mathbb{Q}})D(\lambda^{\mathbb{Q}})^{-1}C(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_X) \\
K_1^Q &= D(\lambda^{\mathbb{Q}})J(\lambda^{\mathbb{Q}})D(\lambda^{\mathbb{Q}})^{-1} \\
\rho_0 &= r_\infty^{\mathbb{Q}} - \rho_1 D(\lambda^{\mathbb{Q}})^{-1}C(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_X) \\
\rho_1 &= \left(D(\lambda_t^{\mathbb{Q}})^{-1}\right)' 1 \\
\Sigma_{\mathcal{P}} &= D(\lambda^{\mathbb{Q}})\Sigma_X D(\lambda^{\mathbb{Q}})'
\end{aligned}
$$

This rotation will preserve the affine structure of yields:

$$
y_{m,t} = A_X\left(r_{t,\infty}^{\mathbb{Q}}, \lambda_t^{\mathbb{Q}}, \Sigma_X\right) + B_X\left(\lambda_t^{\mathbb{Q}}\right)\mathcal{P}_t
$$

The Riccati equations for the model with factor portfiolios will therefore be:

$$
\begin{aligned}
A_{m+1} - A_m &= \left(K_0^{\mathbb{Q}}\right)' B_m + \frac{1}{2}B_m'\Sigma_{\mathcal{P}}B_m - \rho_0 \\
B_{m+1} - B_m &= \left(K_1^{\mathbb{Q}}\right)' B_m - \rho_1
\end{aligned}
$$

C Comparing Recursive and Bayesian Learning

This appendix derives the log likelihood function when assumed that $\psi_t^{\mathbb{P}}$ drifts according to (18) taking $(\Theta^{\mathbb{Q}}, \Sigma_O)$ as known and gives conditions under which the posterior mean μ_t will satisfy the system of equations in (21) - (22). We begin by decomposing the log likelihood into a $\mathbb{P}$ and $\mathbb{Q}$ part

$$
-2\log L = -2\log L^{\mathbb{Q}}(\theta^{\mathbb{Q}}, \Sigma) - 2\log L^{\mathbb{P}}(\Sigma, \{Q_t\})
$$

where the notation $\{Q_t\}$ with abuse of notation indicates the set of free parameters in a given specification of the covariance matrix of the instantaneous parameter state innovations. $\log L^{\mathbb{Q}}$ denotes the part of the likelihood function associated with pricing errors and $\log L^{\mathbb{P}}$ the likelihood function of the dynamic evolution of Z_t. Conditional on $(\Theta^{\mathbb{Q}}, \Sigma_Z)$ the solution to the Kalman filter recursively updates the posterior mean, posterior variance, and forecast variance according to

$$
\mu_t = \mu_{t-1} + P_{t-1}x_{t-1}'\Omega_t^{-1}(y_t - x_{t-1}\mu_{t-1}) \tag{26}
$$

$$
P_t = P_{t-1} + Q_t - P_{t-1}x_{t-1}'\Omega_t^{-1}x_{t-1}P_{t-1} \tag{27}
$$

$$
\Omega_t = x_{t-1}P_{t-1}x_{t-1}' + \Sigma \tag{28}
$$

with log likelihood function given by

$$
-2\log L^{\mathbb{P}} = TN\log(2\pi) + \sum_{t=1}^{T}\log|\Omega_t| + \sum_{t=1}^{T}(y_t - x_{t-1}\mu_{t-1})'\Omega_t^{-1}(y_t - x_{t-1}\mu_{t-1}) \tag{29}
$$

Reworking the first equation gives[26]

$$\mu_t = \mu_{t-1} + (P_t - Q_t)x'_{t-1}\Sigma^{-1}(y_t - x_{t-1}\mu_{t-1})$$

which is (21) for $R_t = (P_t - Q_t)^{-1}$. Equation (27) can then be rewritten into[27]

$$(P_t - Q_t)^{-1} = P_{t-1}^{-1} + x'_{t-1}\Sigma^{-1}x_{t-1} \tag{30}$$

which is (20). The condition for which equation (30) reduces to (22) and therefore is an ALS estimator is $P_{t-1}^{-1}Q_{t-1} = (1-\gamma_t)\cdot I$ for a sequence of scalars $0 < \gamma_t \leq 1$. By reworking (27) it is seen that this condition can be satisfied by choosing

$$Q_t = (\frac{1}{\gamma_{t+1}} - 1)(P_{t-1} - P_{t-1}x'_{t-1}\Omega_t^{-1}x_{t-1}P_{t-1})$$

We can summarize this case as

$$R_t\mu_t = \gamma_t R_{t-1}\mu_{t-1} + x'_{t-1}\Sigma^{-1}y_t \tag{31}$$
$$R_t = \gamma_t R_{t-1} + x'_{t-1}\Sigma^{-1}x_{t-1} \tag{32}$$
$$P_t = \frac{1}{\gamma_{t+1}}R_t^{-1} \tag{33}$$
$$\Omega_t = x_{t-1}P_{t-1}x'_{t-1} + \Sigma \tag{34}$$

with log likelihood function given by (29).

D The Structure of the Particle Filtering Algorithms for the Models with Bayesian Learning

We now turn to the discussion of estimation and results for each model separately. We start from the model with fixed unknown parameters The box below illustrates the estimation recursion for particle i.

If we fix parameter values, the DTSM presented in this model reduces to the full information Gaussian DTSM. Thus, each particle corresponds to an affine term structure model that can be solved using the standard recursions. The entire model is just a mixture of standard GDTSM, where weights correspond to particle weights. This is a very attractive feature of the model; since in each particle pricing is solved in closed form, the computational burden in the algorithm is massively reduced and the entire recursion can be run extremely fast.

Particle Filtering Algorithm:

[26] Substitute (28) into (27) and the resulting equation into (26).

[27] By substituting (28) into (27), plugging the resulting equation back into (27), and multiplying by $(P_t - Q_t)^{-1}$ from the left and P_{t-1}^{-1} from the right.

1 Generate the index $z(i,t-1) \sim Multi_N\left(\{w_{t-1}, \left(\Sigma_{\mathcal{P}}, \Sigma_u, \theta^{\mathbb{Q}}_{t-1}, K_{Z,t-1}\right)^{(i)}\}_{i=1}^{N}\right)$, where $Multi_N$ is a multinomial distribution.

2 Generate $\theta_t^{(i),\mathbb{Q}} \sim p\left(\theta_t^{\mathbb{Q}} | \left(\theta^{\mathbb{Q}}\right)^{z(i,t-1)}, \mathcal{P}_t, Y_t^o\right)$:

- We draw $\theta_t^{(i),\mathbb{Q}} \sim \mathcal{B}\left(a^(i), b^(i)\right)$ for $k \in \{1, ..., K\}$.
- We set $a^(i)$ and $b^(i)$ so that the mean of the beta distribution is equal to a mix of the weighted mean of $\theta^{\mathbb{Q}}_{t-1}$ and $\theta^{z(i,t),\mathbb{Q}}_{t-1}$. Thus we calculate:

$$\mu_{t-1}^{(i),\theta} = (1-\alpha)\sum_{j=1}^{M} w_{t-1}^{(j)}\theta_{t-1}^{(j),\mathbb{Q}} + \alpha\theta_{t-1}^{z(i,t),\mathbb{Q}}$$

  And set:

$$a^{(i)} = \mu_{t-1}^{(i),\theta}\phi$$
$$b^{(i)} = \left(1-\mu_{t-1}^{(i),\theta}\right)(1+\phi)$$

  The parameters α and ϕ is available for calibration purposes; we set $\phi = 1000$ and $\alpha = 0.1$.

- We can solve yields pricing equation and obtain $A^{(i)}$ and $B^{(i)}$.

3 Generate $K_t^{(i)} \sim p\left(K_t | \left(K_{t-1}, \mathcal{P}_t, \mathcal{P}_{t-1}, \Sigma_{\mathcal{P}}\right)^{z(i,t)}\right)$, which is a vectorized representation of the intercepts and the mean reversion matrix in the model. We can get conditional means and covariance matrix of the parameters as:

- $R = C_{t-1}^{z(i,t-1),K}$
- $X_{t-1} = I \otimes [1 \quad \mathcal{P}_{t-1}]$
- $\mu_t^{(i),K} = \mu_{t-1}^{z(i,t-1),K} + RX'_{t-1}\left(X_{t-1}RX'_{t-1} + I \otimes SS_{\mathcal{P}}^{z(i,t-1)}\right)^{-1}\left(\mathcal{P}_t - \mu_{t-1}^{z(i,t-1),K}X_{t-1}\right)$
- $C_t^{(i),K} = R - RX'_{t-1}\left(X_{t-1}RX'_{t-1} + SS_{\mathcal{P}}^{z(i,t-1)}\right)^{-1}X_{t-1}R'$
- Draw: $K_t^{(i)} \sim N\left(\mu_t^{(i),K}, C_t^{(i),K}\right)$

4 Compute sample statistics $s_{t+1} = \mathcal{S}\left(s_t^{z(i,t)}, \theta_{t+1}^{(i),\mathbb{Q}}, Y_{t+1}^o, \mathcal{P}_{t+1}\right)$.

- $e_{\mathcal{P},t} = \mathcal{P}_t - K_t^{(i)} \mathcal{P}_{t-1}$
- $S_{\mathcal{P},t} = S_{\mathcal{P},t-1} + \frac{1}{2} e_{\mathcal{P},t} e_{\mathcal{P},t}^{'}$
- $u_t = Y_t^o - A^{(i)} - B^{(i)} \mathcal{P}_t$
- $S_{u,t} = S_{u,t-1} + \frac{1}{2} u_t u_t^{'}$

5 Generate $\left(\Sigma_{\mathcal{P}}^{(i)}\right)^{-1} \sim \mathcal{W}\left(v_{\mathcal{P},t}, S_{\mathcal{P},t}\right)$, with $v_{\mathcal{P},t} = v_{\mathcal{P},t-1} + \frac{1}{2}$.

6 Generate $\left(\Sigma_u^{(i)}\right)^{-1} \sim \mathcal{W}\left(v_{u,t}, S_{u,t}\right)$, with $v_{u,t} = v_{u,t-1} + \frac{1}{2}$.

7 Compute particle weights $w_t^{(i)} = \frac{p\left(\mathcal{P}_{t+1}, Y_{t+1}^o | \mathcal{P}_t, Y_t^o, (\Sigma_{\mathcal{P}}, \Sigma_u, \theta_{t+1}^{\mathbb{Q}}, K_{t+1})^{(i)}\right)}{\sum_{j=1}^N p\left(\mathcal{P}_{t+1}, Y_{t+1}^o | \mathcal{P}_t, Y_t^o, (\Sigma_{\mathcal{P}}, \Sigma_u, \theta_{t+1}^{\mathbb{Q}}, K_{t+1})^{(j)}\right)}$

Particle Filtering Algorithm (Drifting Mean Parameters):

1 Generate the index $z(i,t-1) \sim Multi_N\left(\{w_{t-1}, \left(\Sigma_{\mathcal{P}}, \Sigma_u, \Sigma_w, \theta_{t-1}^{\mathbb{Q}}, K_{t-1}\right)^{(i)}\}_{i=1}^N\right)$, where $Multi_N$ is a multinomial distribution.

2 Identical to step 2 in the basic algorithm.

3 Generate $K_t^{(i)} \sim p\left(K_t | \left(K_{t-1}, \Sigma_w\right)^{z(i,t)}\right)$, which consists of drawing $K_t^{(i)} \sim N\left(K_{t-1}^{z(i,t)}, \Sigma_w^{z(i,t)}\right)$.

4 Compute sample statistics $s_{t+1} = \mathcal{S}\left(s_t^{z(i,t)}, \theta_{t+1}^{(i),\mathbb{Q}}, Y_{t+1}^o, \mathcal{P}_{t+1}\right)$.

- $e_{\mathcal{P},t} = \mathcal{P}_t - K_t^{(i)} \mathcal{P}_{t-1}$
- $S_{\mathcal{P},t} = S_{\mathcal{P},t-1} + \frac{1}{2} e_{\mathcal{P},t} e_{\mathcal{P},t}^{'}$
- $u_t = Y_t^o - A^{(i)} - B^{(i)} \mathcal{P}_t$
- $S_{u,t} = S_{u,t-1} + \frac{1}{2} u_t u_t^{'}$

- $w_t = K_t^{(i)} - K_{t-1}^{z(i,t)}$
- $S_{w,t} = S_{w,t-1} + \frac{1}{2} w_t w_t'$

5 Generate $\left(\Sigma_{\mathcal{P}}^{(i)}\right)^{-1} \sim \mathcal{W}\left(v_{\mathcal{P},t}, S_{\mathcal{P},t}\right)$, with $v_{\mathcal{P},t} = v_{\mathcal{P},t-1} + \frac{1}{2}$.

6 Generate $\left(\Sigma_{u}^{(i)}\right)^{-1} \sim \mathcal{W}\left(v_{u,t}, S_{u,t}\right)$, with $v_{u,t} = v_{u,t-1} + \frac{1}{2}$.

7 Generate $\left(\Sigma_{w}^{(i)}\right)^{-1} \sim \mathcal{W}\left(v_{w,t}, S_{w,t}\right)$, with $v_{w,t} = v_{w,t-1} + \frac{1}{2}$.

8 Compute particle weights $w_t^{(i)} = \frac{p\left(\mathcal{P}_{t+1}, Y_{t+1}^o | \mathcal{P}_t, Y_t^o, (\Sigma_{\mathcal{P}}, \Sigma_u, \Sigma_w, \theta_{t+1}^{\mathbb{Q}}, K_{t+1})^{(i)}\right)}{\sum_{j=1}^N p\left(\mathcal{P}_{t+1}, Y_{t+1}^o | \mathcal{P}_t, Y_t^o, (\Sigma_{\mathcal{P}}, \Sigma_u, \theta_{t+1}^{\mathbb{Q}}, K_{t+1})^{(j)}\right)}$

E Bond Pricing with Deterministically Drifting Coefficients

We now describe the particle filter algorithm that we use to solve the joint filtering and learning problem. As discussed in the body of the article, in a model with drifting coefficients and parameter uncertainty, it is not possible to preserve the affine structure of yields. Nonetheless, the model is still affine *conditionally on the realizations of the parameters*. In particular, consider the *conditional* zero coupon bond pricing recursion, for maturities between 1 and M periods:

$$
\begin{aligned}
D_{1,t} &= E^{\mathbb{Q}}\left[\exp(-r_t) | \lambda_t^{\mathbb{Q}}, r_{\infty,t}^{\mathbb{Q}}, \Sigma_{X,t}\right] \\
D_{2,t} &= E^{\mathbb{Q}}\left[D_{1,t+1} \exp(-r_t) | \left(\lambda_{t+i}^{\mathbb{Q}}, r_{\infty,t+i}^{\mathbb{Q}}, \Sigma_{X,t+i}\right)_{i=0}^{1}\right] \\
&\dots \\
D_{M,t} &= E^{\mathbb{Q}}\left[D_{M-1,t+1} \exp(-r_t) | \left(\lambda_{t+i}^{\mathbb{Q}}, r_{\infty,t+i}^{\mathbb{Q}}, \Sigma_{X,t+i}\right)_{i=1}^{M-1}\right]
\end{aligned}
$$

If we fix a deterministic path of the coefficients: $\left(\lambda_{t+i}^{\mathbb{Q}}, r_{\infty,t+i}^{\mathbb{Q}}, \Sigma_{\mathcal{P},t+i}\right)_{i=0}^{M}$, we still find that model implied yields are an affine function of the pricing factors:

$$
y_{t,m} = \tilde{A}_{X,m} + \tilde{B}_{X,m} X_t
$$

Where $\tilde{A}_{X,m}$ and $\tilde{B}_{X,m}$ are given by the recursions:

$$\tilde{A}_{X,m} - \tilde{A}_{X,m-1} = \frac{1}{2}\tilde{B}'_{X,m-1}\Sigma_{X,t}\tilde{B}_{X,m-1} - r^{\mathbb{Q}}_{\infty,t}$$
$$\tilde{B}_{X,m} - \tilde{B}_{X,m-1} = J\left(\lambda^{\mathbb{Q}}_t\right)'\tilde{B}_{X,m-1}$$

Thus, the coefficients in the Riccati equations will depend on the entire future path of the coefficients determining the risk-adjusted dynamics of the factors. Then, we again assume that pricing factors can be exactly rotated into portfolios. It follows that:

$$\mathcal{P}_t = WA_{y,X} + WB_{y,X}X_t$$
$$\mathcal{P}_t = C_t + D_tX_t$$

Then, the conditional Q expectations of the factor portfolios can be obtained as:

$$E^{\mathbb{Q}}\left[\mathcal{P}_{t+1}|\left(\lambda^{\mathbb{Q}}_{t+i}, r^{\mathbb{Q}}_{\infty,t+i}, \Sigma_{\mathcal{P},t+i}\right)^{M+1}_{i=0}\right] = \tilde{C}_{t+1} + \tilde{D}_{t+1}E^{\mathbb{Q}}_t\left[X_{t+1}|\left(\lambda^{\mathbb{Q}}_{t+i}, r^{\mathbb{Q}}_{\infty,t+i}, \Sigma_{\mathcal{P},t+i}\right)^{M+1}_{i=0}\right]$$
$$= \tilde{C}_{t+1} + \tilde{D}_{t+1}J\left(\lambda^{\mathbb{Q}}_t\right)\tilde{D}^{-1}_t\left(\mathcal{P}_t - \tilde{C}_t\right)$$

While their conditional variances will be:

$$Var^{\mathbb{Q}}\left[\mathcal{P}_{t+1}|\left(\lambda^{\mathbb{Q}}_{t+i}, r^{\mathbb{Q}}_{\infty,t+i}, \Sigma_{\mathcal{P},t+i}\right)^{M+1}_{i=0}\right] = \tilde{D}_{t+1}Var^{\mathbb{Q}}\left[X_{t+1}|\left(\lambda^{\mathbb{Q}}_{t+i}, r^{\mathbb{Q}}_{\infty,t+i}, \Sigma_{\mathcal{P},t+i}\right)^{M+1}_{i=0}\right]\tilde{D}_{t+1}$$
$$= \tilde{D}_{t+1}\Sigma_{X,t}\tilde{D}_{t+1}$$

We can rewrite the model in term of the observable portfolios:

$$\mathcal{P}_{t+1} = K^{\mathbb{Q}}_{0,t} + K^{\mathbb{Q}}_{1,t}\mathcal{P}_t + \Sigma^{1/2}_{\mathcal{P}}e_{\mathcal{P},t+1}$$
$$r_t = \rho_{0,t} + \rho_{1,t}\mathcal{P}_t$$

References

Altavilla, C., R. Giacomini, and G. Ragusa, 2014, Anchoring the Yield Curve Using Survey Expectations, Working paper, Working Paper No 1632, European Central Bank.

Ang, A., and G. Bekaert, 2002, Regime Switches in Interest Rates, *Journal of Business and Economic Statistics* 20, 163–182.

Barillas, F., L. Hansen, and T. Sargent, 2009, Doubts or Variability?, *Journal of Economic Theory* 144, 2388–2419.

Bauer, M., G. Rudebusch, and J. Wu, 2013, Comment on “Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset”, *American Economic Review.*

Buraschi, A., and P. Whelan, 2012, Term Structure Models with Differences in Beliefs, Working paper, Imperial College.

Cieslak, A., and P. Povala, 2014, Expecting the Fed, Working paper, Northwestern University.

Cochrane, J., and M. Piazzesi, 2005, Bond Risk Premia, *American Economic Review* 95, 138–160.

Cogley, T., and T. Sargent, 2005, Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S., *Review of Economics Dynamics* 8, 262–302.

Cogley, T., and T. Sargent, 2008, Anticipated Utility and Rational Expectations as Approximations of Bayesian Decision-Making, *International Economic Review* 49, 185–221.

Collin-Dufresne, P., M. Johannes, and L. Lochstoer, 2013, Parameter Learning in General Equilibrium: The Asset Pricing Implications, Working paper, Columbia University.

Dai, Q., and K. Singleton, 2000, Specification Analysis of Affine Term Structure Models, *Journal of Finance* 55, 1943–1978.

Dai, Q., and K. Singleton, 2002, Expectations Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure, *Journal of Financial Economics* 63, 415–441.

Dai, Q., K. Singleton, and W. Yang, 2007, Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields, *Review of Financial Studies* 20, 1669–1706.

Dewachter, H., and M. Lyrio, 2008, Learning, Macroeconomic Dynamics and the Term Structure of Interest Rates, in *Asset Prices and Monetary Policy* (National Bureau of Economic Research,).

Diebold, F., and R. Mariano, 1995, Comparing Predictive Accuracy, *Journal of Business and Economic Statistics* 13, 253–263.

Duarte, J., 2008, The Causal Effect of Mortgage Refinancing on Interest-Rate Volatility: Empirical Evidence and Theoretical Implications, *Review of Financial Studies* 21, 1689–1731.

Duffee, G., 2001, Term Premia and Interest Rates Forecasts in Affine Models, forthcoming, Journal of Finance.

Duffee, G., 2002, Term Premia and Interest Rate Forecasts in Affine Models, *Journal of Finance* 57, 405–443.

Duffee, G., 2010, Sharpe Ratios in Term Structure Models, Working paper, Johns Hopkins University.

Duffee, G., 2011, Forecasting with the Term Structure: the Role of No-Arbitrage, Working paper, Johns Hopkins University.

Feldhutter, P., L. Larsen, C. Munk, and A. Trolle, 2012, Keep It Simple: Dynamic Bond Portfolios Under Parameter Uncertainty, Working paper, London Business School.

Fleming, M. J., and E. Remolona, 1999, The Term Structure of Announcement Effects, FRB New York Staff Report No. 76.

Hansen, L., 2007, Beliefs, Doubts, and Learning: Valuing Macroeconomic Risk, *American Economic Review* 97, 1 – 30.

Hansen, L., and S. Richard, 1987, The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models, *Econometrica* 55, 587–613.

Harvey, D., S. Leybourne, and P. Newbold, 1997, Testing the Equality of Prediction Mean-Squared Errors, *International Journal of Forecasting* 13, 281–291.

Hu, X., J. Pan, and J. Wang, 2013, Noise as Information for Illiquidity, *Journal of Finance* 68, 2223–2772.

Jardet, C., A. Monfort, and F. Pegoraro, 2012, No-Arbitrage Near-Cointegrated VAR(p) Term Structure Models, Term Premia and GDP Growth, Working paper, Banque de France.

Joslin, S., and A. Le, 2013, Interest Rate Volatility and No Arbitrage Term Structure Models, Working paper, University of North Carolina.

Joslin, S., A. Le, and K. Singleton, 2013, Gaussian Macro-Finance Term Structure Models with Lags, *Journal of Financial Econometrics* 11, 581–609.

Joslin, S., M. Priebsch, and K. Singleton, 2013, Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks, Working paper, forthcoming, *Journal of Finance.*

Joslin, S., K. Singleton, and H. Zhu, 2011, A New Perspective on Gaussian Dynamic Term Structure Models, *Review of Financial Studies* 24, 926–970.

Jouini, E., and C. Napp, 2007, Consensus Consumer and Intertemporal Asset Pricing with Heterogeneous Beliefs, *Review of Economic Studies* 74, 1149–1174.

Kim, D., and A. Orphanides, 2012, Term Structure Estimation with Survey Data on Interest Rate Forecasts, *Journal of Financial and Quantitative Analysis* 47, 241–272.

Krishnamurthy, A., and A. Vissing-Jorgensen, 2011, The Effects of Quantitative Easing on Long-term Interest Rates, *Bookings Papers on Economic Activity* Fall 2011.

Laubach, T., R. Tetlow, and J. Williams, 2007, Learning and the Role of Macroeconomic Factors in the Term Structure of Interest Rates, Working paper, Board of Governors of the Federal Reserve System.

Le, A., and K. Singleton, 2012, A Robust Analysis of the Risk-Structure of Equilibrium Term Structures of Bond Yields, Working paper, University of North Carolina.

Le, A., K Singleton, and J. Dai, 2010, Discrete-Time Affine^Q Term Structure Models with Generalized Market Prices of Risk, *Review of Financial Studies* 23, 2184–2227.

Litterman, R., and J. Scheinkman, 1991, Common Factors Affecting Bond Returns, *Journal of Fixed Income* 1, 54–61.

McCulloch, J., 2007, The Kalman Foundations of Adaptive Least Squares, With Applications to U.S. Inflation, Working paper, Ohio State University.

Piazzesi, M., 2005, Bond Yields and the Federal Reserve, *Journal of Political Economy* 113, 311–344.

Piazzesi, M., J. Salomao, and M. Schneider, 2013, Trend and Cycle in Bond Premia, Working paper, Stanford University.

Rudebusch, G., and T. Wu, 2008, A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy, *Economic Journal* 118, 906–926.

Sims, C., and T. Zha, 2006, Were There Regime Changes in Monetary Policy?, *American Economic Review* 96, 54–81.

Swanson, E., 2006, Have Increases in Federal Reserve Transparency Improved Private Sector Interest Rate Forecasts?, *Journal of Money, Credit, and Banking* 38, 791–819.

Ulrich, M., 2013, Inflation Ambiguity and the Term Structure of U.S. Government Bonds, *Journal of Monetary Economics* 60, 295–309.

Wright, J., 2011, Term Premiums and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset, *American Economic Review* 101, 1514–34.

Xiong, W., and H. Yan, 2009, Heterogeneous Expectations and Bond Markets, *Review of Financial Studies.*